Dec. 13

JUL 28 1954

Social Science Library

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Missed Signals in a Sensorimotor Skill: R. CONRAD 1

Psychomotor Response Acquisition and Transfer as a Function of Control-Indicator Relationships: J. A. ADAMS 10

Multiple versus Single Problem Training in Human Problem Solving: J. A. ADAMS 15

Incentive Conditions Which Influence Visual Exploration: R. A. BUTLER 19

Learning Variables in the Judgment of Single Stimuli: A. PARDUCCI 24

The Effects of a Pre-Extinction Procedure on the Extinction of Place and Response Performance in a T Maze: D. P. SCHARLOCK 31

Motivational Properties of Frustration: II. Frustration Drive Stimulus and Frustration Reduction in Selective Learning: A. AMSEL AND J. S. WARD 37

Studies in Incidental Learning: I. The Effects of Crowding and Isolation: L. POSTMAN AND L. W. PHILLIPS 48

Maximum Rates of Form Perception and the Alpha Rhythm: An Investigation and Test of Current Nerve Net Theory: O. D. MURPHREE 57

An Attempt to Obtain Pupillary Conditioning with Infrared Photography: F. A. YOUNG .. 62

The Reliability of Measurements of Human Dark Adaptation: F. A. MOTE, G. E. BRIGGS, AND K. M. MICHELS 69

Effect of Number of Alternatives and Set on the Visual Discrimination of Numerals: G. K. KRULEE, J. E. PODELL, AND P. G. RONCO 75

American Psychological Association

Vol. 48 No. 1 July 1954

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1954 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright, 1954, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 48, NO. 1 JULY, 1954

MISSED SIGNALS IN A SENSORIMOTOR SKILL [1]

R. CONRAD

Applied Psychology Research Unit, Cambridge, England

The experiment reported is one of a series that aims at an understanding of the conditions for the maintenance of skilled behavior. In particular, attention is focused on changes in performance that occur when systematic variations are made in the rate of change of a continuously changing visual display.

Previous work (3) has shown how in a sensorimotor skill in which *S* was required to deal with four concurrent series of signals, response was to a large extent adapted to the temporal exigencies of the moment. The effect was that responses did not necessarily coincide in time exactly with the signals that initiated them. About half were made a little too late, and half a little too early. This was not a chance effect, but was shown to depend on the time since the previous response, and also on the time to the signal next to be responded to.

An earlier investigation (2) had demonstrated that some signals were not reponded to at all. These omissions increased disproportionately when the mean signal speed was increased. The present paper attempts to determine the display factors underlying the occurrence of signals to which *S* fails to respond.

METHOD

The apparatus, which has been fully described by Conrad (2), consisted of a display of four pointers, each moving at slightly different speeds around four dials, each of which carried six equidistant marks around the perimeters. The speed of the pointers and hence the mean signal speed could be controlled by *E*. The *S*'s task was to respond by tapping a Morse key, with either hand, whenever a pointer coincided with a perimeter mark. This response did not affect the movement of the pointers, but its moment of occurrence was recorded on moving paper, as was the actual moment of coincidence, which was the signal for action. The *S* could always see the pointers and the target marks on the dials, and the skill lay in the ability to determine which was the next pointer to merit attention in order to make an accurate response.

Since the relative time intervals between successive signals on each of the four dials were 1, 1.09, 1.40, 1.67, it will be recognized that, taking the display as a whole, signals occurred neither in a regular spatial order nor at equal intervals of time. A mathematical analysis of the character of the distribution of this interval has been made by Cox and Smith (4).

Six naval ratings served as *S*s. They were given a 15-min. practice period, morning and afternoon, for two days, during which all the subsequent speeds were introduced. Three similar experimental sessions then followed during which each *S* did 3, 2, and 2 tests, respectively,

[1] The work was carried out under the general direction of Sir Frederic Bartlett C.B.E., F.R.S., and Dr. N. H. Mackworth to whom the author is deeply indebted.

TABLE 1

EFFECT OF SPEED CHANGES ON PERFORMANCE

Measure	Speed (signals/min.)						
	40	60	80	100	120	140	160
Omissions/minute	.3	3.7	12.2	28.2	41.3	63.1	80.0
Responses/minute	39.7	56.3	67.8	71.8	78.7	76.9	80.0
Accuracy (sec.)	.26	.26	.25	.24	.25	.24	.24

each of 5-min. duration. The following seven conditions of speed were given in random order: 40, 60, 80, 100, 120, 140, 160 signals/min. Each experimental session was preceded by a 3-min. "warming up."

RESULTS AND DISCUSSION

The problem of omissions.—Table 1 gives the general results. It became clear merely from looking at the records, that there were several probable reasons that would account for some of the omissions, but that none of them occurred consistently on the records of all *S*s at all speeds. Principally these factors were: (*a*) attentional lapses occurring at very slow speeds, involving one signal; (*b*) blocks at fast speeds, which appeared to be blocks of the kind described by Bills (1), and lasted for periods of about 1–3 sec. during which all signals were missed; and (*c*) narrowing of the attentional span at fast speeds. With respect to *c*, one or two *S*s appeared to find it difficult to attend to events from four sources concurrently. This showed as stretches on their record of perhaps 30-sec. duration when signals on one dial were completely neglected. The dial neglected usually changed continually during this period.

The general description of these three broad possibilities did not seem to provide a satisfactory explanation of the omission data. Essentially, it was desirable to isolate one or more features of the display with which omissions tended to be associated more often than would be expected by chance.

Table 1 rules out the possibility that responses were made entirely at random, and that therefore omissions were random. The accuracy data alone indicate that responses were associated in time with signals. It is possible to calculate what the mean size of error would be if the position of a response was independent of the temporal position of any signal. In such a case the faster the signal speed, the more likely was a response to be close to a signal. If responses were distributed among the four response keys in numbers proportionate to the number of signals occurring on each of the four dials, then the mean error would be .43 sec. at 160 signals/min., which is nearly double that actually measured. At all other speeds the calculated mean error by chance would be more than double that observed.

A hypothesis that deserves more careful investigation is that signals were *selected* for response without regard to the position of other signals or responses, but once selected, were dealt with fairly accurately. The hypothesis obviously cannot be tested for slow speeds since there are too few omissions; that is, nearly all the signals were selected. But it can be tested, and is most likely to hold, at the fastest speed.

We know from Table 1 that at a signal speed of 160 signals/min., 80

responses/min. were made. If 80 signals out of the 160 were selected without regard to the position of other events, then the frequency distribution of the interval between these 80 signals responded to would closely resemble that for a signal speed of 80 signals/min. The difference will be in the direction of the exponential. Data are available for 736 responses at the fastest speed. The observed distribution of the interval between the signals responded to is given in Table 2 together with the expected distribution.

As measured by chi square, the observed distribution departs significantly from the expected, and it is evident that the observed distribution contains too few short intervals and too many long ones.

The observed distribution of the interval between signals responded to also permits the rejection of a hypothesis that signals were selected for response at approximately equal intervals of time. Equally unsupported, of course, is the possibility that responses were made at regular intervals. This is essentially the same as the previously rejected hypothesis that responses were random in time relative to the position of the signals.

Signal-response sequence.—With scant support for the immediately obvious hypotheses by which the occurrence of omissions might be explained, a more detailed examination of the temporal structure of the signal-response series was made.

TABLE 2

FREQUENCY DISTRIBUTION OF INTERVAL BETWEEN SUCCESSIVE SIGNALS RESPONDED TO AT SPEED 160 SIGNALS/MIN. AND DISTRIBUTION EXPECTED IF SIGNALS WERE SELECTED FOR RESPONSE AT RANDOM

Frequency	Interval (.25 sec.)						
	1	2	3	4	5	5+	Total
Observed	120	116	119	156	113	112	736
Expected	18	14	15	9	10	14	80

The method used to reconstruct the performance records so as to permit the necessary sequence analysis to be made has been previously described (3). Briefly, it involved linking all the responses to the signals to which they were made, and then setting them on a single time base: this gave a series of signals spaced according to the time intervals between them, irrespective of the dial on which they occurred, and associated with it a series of responses. On the same time base appeared those signals to which no response was made—the omissions. Any signal not responded to stands in some simple temporal relationship with many other events, e.g.: (*a*) the previous signal whether responded to or not; (*b*) the next signal whether responded to or not; (*c*) the previous signal responded to; (*d*) the next signal responded to; (*e*) the previous response; and (*f*) the next response. Each of these relationships can be used as a basis for a hypothesis. But in view of the large number of possible hypotheses that could be set up and the labor involved in testing each of them, it was highly desirable that the least obviously useful should be discarded.

Accordingly, pilot analyses were made of small samples of the recorded data. All of the six relationships mentioned above were examined. In each case, the distribution of the interval between signals responded to and the particular event was compared with that of the interval between signals omitted and the same event. In no case was the difference significant, and thus no hypothesis succeeded.

It became evident, however, during this preliminary analysis that although omissions did not tend to occur either consistently shortly before or consistently shortly after the events listed, they did seem in certain cases to fall in one or the other of these categories. Omissions appeared frequently to be *close* to another signal, or *close* to a response.

The relationship between omissions and proximate signals, even if established, is of little theoretical value in the present context, since it contains minimal information about *S*'s behavior. In fact, it can be misleading. In many instances there would be a much more significant event between the two mentioned, that is, a response. Without our invoking an unsatisfactory number of assumptions, no assessment of the psychological significance of the relationship between omissions and the nearest signal could be made. Attention was therefore centered on testing the hypothesis that omissions occurred closer to the nearest response than would be expected by chance.

To examine this hypothesis a 2-min. sample taken from the middle of the test period of each *S* was analyzed for the 100-, 120-, 140-, and 160-signals/min. speeds. For the speeds, 40, 60, and 80 signals/min., the whole 5-min. test period was analyzed, since there were relatively few omissions in these cases. Using a unit of interval size of .25 sec., the time interval from

TABLE 3

INTERVAL BETWEEN EVERY SIGNAL AND THE NEAREST RESPONSE—THE NUMBER OF TIMES, FOR EACH SIZE OF INTERVAL, THAT THE SIGNAL WAS MISSED AND RESPONDED TO

Measure*	Interval Size (.25 sec.)										Total	Speed
	1	2	3	4	5	6	7	8	9	10		
O	7	2	1	0	0	0	0	0	0	0	10	40
R	265	194	225	182	120	70	62	40	28	4	1,190	spm
T	272	196	226	182	120	70	62	40	28	4	1,200	
O	56	33	11	8	2	1	1				112	60
R	492	418	314	247	133	60	24				1,688	spm
T	548	451	325	255	135	61	25				1,800	
O	172	100	62	24	6	2	0				366	80
R	615	629	453	242	59	25	11				2,034	spm
T	787	729	515	266	65	27	11				2,400	
O	172	98	38	17	8	3	2				338	100
R	287	271	167	81	32	18	6				862	spm
T	459	369	205	98	40	21	8				1,200	
O	271	136	61	20	5	3					496	120
R	352	287	190	79	27	9					944	spm
T	623	423	251	99	32	12					1,440	
O	422	240	75	17	3	0					757	140
R	307	293	200	77	38	8					923	spm
T	729	533	275	94	41	8					1,680	
O	544	287	93	31	4	0					959	160
R	311	319	204	90	33	4					961	spm
T	855	606	297	121	37	4					1,920	

* O is the number of signals missed; R is the number of signals responded to; and T is the total number found of the particular size of interval.

every omission to the nearest response was measured, regardless of whether this response came before or after the signal in question. In fact, the nearest response occurred before and after omissions in approximately equal numbers. Similarly, the time was also measured between every signal responded to and the nearest response. The frequency distributions are given in Table 3.

Position of omissions.—It is evident from Table 3 that at all speeds omissions did in fact occur close to responses. At every speed about three-quarters of the omissions occurred within .5 sec. of a response, and about one-half of them within .25 sec. The problem is whether or not they occurred close to a response more often than would be expected by chance. The probability that an omission will occur close to a response depends partly on the signal speed and partly on the response rate. In the first case, the more frequently the signals occur, the greater the statistical probability that one of them will fall near to a response. In the second case, the more frequently responses are made, the greater the probability that a missed signal will occur close to one of them. The combined probability cannot be calculated simply, because even though the response rates are known as well as the signal speeds, the responses are not necessarily accurately timed, and the size of the time error determines the size of the interval between signal and response.

If chance factors alone determined whether or not a signal was responded to, then the distribution of every 100 intervals between omissions and nearest response would be the same as that of every 100 intervals between signals responded to and the nearest response. If, however, more omissions occur close to responses than would be expected by chance, then these two distributions will differ. By use of the data in Table 3, these distributions are set up in Table 4, which compares the proportion of omissions with the proportion of signals responded to in five different intervals from the nearest response. For convenience the former is regarded as the observed (Obs.) and the latter as the expected (Exp.) distribution. For intervals longer than 1.25 sec. the relationship between the expected and observed proportions remained unchanged—there were always more expected than observed.

For the full data given in Table 3, tested by chi square, the two distributions are found to be significantly different at all speeds. The difference lies in the fact that proportionately more omissions than would be expected by chance occur within .25 sec. of the nearest response, and proportionately fewer when the interval is greater than

TABLE 4

DISTRIBUTION OF INTERVAL FROM SIGNALS RESPONDED TO (EXP.) AND FROM OMISSIONS (OBS.) TO NEAREST RESPONSE (PROPORTIONS)

Signals/ Min.	Measure	Interval Size (.25 sec.)				
		1	2	3	4	5
40	Exp.	.22	.16	.19	.15	.10
	Obs.	.70	.20	.10	.00	.00
60	Exp.	.29	.25	.19	.15	.08
	Obs.	.50	.30	.10	.07	.03
80	Exp.	.30	.31	.22	.12	.03
	Obs.	.47	.27	.17	.07	.02
100	Exp.	.33	.31	.19	.09	.05
	Obs.	.51	.29	.11	.05	.03
120	Exp.	.37	.30	.20	.08	.04
	Obs.	.55	.27	.12	.04	.01
140	Exp.	.33	.32	.22	.08	.04
	Obs.	.56	.32	.10	.02	.00
160	Exp.	.32	.33	.21	.10	.03
	Obs.	.57	.30	.10	.03	.00

TABLE 5

INTERVAL BETWEEN EVERY SIGNAL AND THE NEAREST RESPONSE: PROPORTION OF SIGNALS MISSED AT DIFFERENT INTERVALS

(Values in parentheses are calculated)

Signals/ Min.	Interval Size (.25 sec.)				
	1	2	3	4	5
40	.03(.02)	.01(0)	0(0)	0(.01)	0
60	.10(.12)	.07(.09)	.03(.06)	.03(.05)	.02
80	.22(.23)	.14(.17)	.12(.11)	.09(.09)	.09
100	.38(.34)	.27(.25)	.19(.17)	.17(.13)	.20
120	.44(.45)	.32(.33)	.24(.22)	.20(.17)	.16
140	.58(.56)	.45(.41)	.27(.27)	.18(.21)	.07
160	.64(.66)	.47(.50)	.31(.33)	.26(.25)	.11

.5 sec. To summarize this point, it can be stated that omissions occur close to responses significantly more often than would be expected by chance. This means that if one signal occurs .25 sec. from a response and another .75 sec. from a response, the former is more likely to be missed than the latter.

The data in Table 3 can also be used to calculate the probability of omission when a signal occurs at a given interval from a response. That is, if a signal occurs .25 sec. from the nearest response, what are the chances that it will be missed or responded to compared with a signal occurring .75 sec. from the nearest response? These probabilities are shown in Table 5 from which, for the sake of simplicity, large intervals have been discarded.

The farther away in time a signal is from the nearest response, the less likely it is to be missed. For example, at 100 signals/min. out of every 100 times that a signal occurred within .25 sec. of a response, it was missed 38 times. But out of every 100 times that a signal occurred between .5 and .75 sec. of a response, it was missed only 19 times. The trend is true at all speeds. It should be noted that this fact—the data in Table 5—does not prove the point made earlier (Table 4) that omissions occurred close to responses. If in fact very few *signals* occurred close to a response, then even if all were missed there would still be very few omissions. Table 4 indicates *where* an omission will occur *if* it occurs. Table 5 indicates *how many* omissions will occur.

It is, however, evident from Table 5 that the probability that a signal is missed depends not only on the time interval to the nearest response, but also on the mean signal speed. For example, at a speed of 40 signals/min., out of every 100 times that a signal occurred within .25 sec. of a response, it was missed 3 times. But at a speed of 160 signals/min. out of every 100 times that a signal occurred within the same time interval of a response, it was missed 64 times. The probability of an omission for a number of intervals is plotted against signal speed in Fig. 1, and it will be seen that the relationship is close to a linear one.

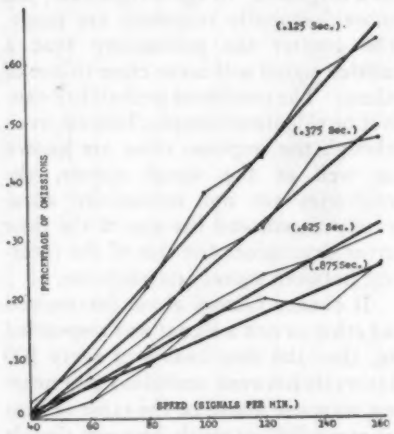


FIG. 1. Proportion of omissions as a function of speed and time interval to nearest response. Mean time to nearest response given in parentheses.

The calculated values are shown in parentheses in Table 5. Although the trends shown represent the mean results from all six *S*s, the individual values show similar trends in all cases.

It appears clear that both before and after a response there is a period of diminished activity. This is not a period of diminished *effector* activity. The position of a response in relation to the nearest response is not being considered here. In any case the task was a two-handed one and there was no bar on the effector side to two simultaneous responses. The change in performance near the time of the response must be due to a change either in the peripheral receptor or the central mechanisms. If effector refractoriness only were being considered, one can see no reason why the response should not be delayed until the effector system was free, in which case early responses would be rare, and the position of omissions would be independent of the position of responses. Neither of these is true. At the same time, the intensity of the diminished activity is related to the signal speed. For any given interval between signal and nearest response the effect is felt more strongly as the mean signal speed increases.

The concept of speed stress.—At first sight the explanation of the considerable speed effect would seem to lie in the detailed temporal changes that occur when the mean speed is increased. The probability that any event will be closely preceded or succeeded by another event will increase as the mean speed increases. If, therefore, the ability to deal with a sequence consisting of a signal and the nearest response to it is affected by further events before or after, then more omissions will be expected per unit number of signals at faster speeds.

This point can be stated hypothetically as follows: if a signal occurs at a given interval from the nearest response, then for a constant mean signal speed, the probability of an omission depends on the time interval to the nearest events before or after any of the other events in the sequence.

There are three ways of examining this: (*a*) to consider every combination of each event of the sequence with every possible event before and after the sequence, either until evidence for the hypothesis is found or until the possible combinations are exhausted; (*b*) to adduce evidence for an alternative hypothesis that might weaken the validity of the present one; and (*c*) to test the hypothesis for the combination most favorable to it, and if it failed, to try to show evidence for an alternative. This third course, a compromise of the other two, was the one adopted.

Initially it was decided, for the sake of simplicity, to consider one signal speed (120 signals/min.) and one size of signal-nearest-response interval (.25 sec.). Measurement was made between the *first* event of the sequence and the *nearest* event before it, and between the *last* event of the sequence

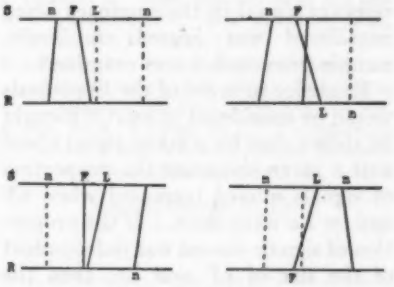


FIG. 2. Examples of the sequence: signal and nearest response, showing the first (F) and last (L) events of the sequence and the nearest event (n) to each. Solid line joins a signal to the response to it. Dotted line indicates position of signal missed.

TABLE 6

PROPORTION OF SIGNALS MISSED CLASSIFIED ACCORDING TO THE TIME INTERVAL BETWEEN THE SEQUENCE AND THE NEAREST EVENTS

Signal speed: 120 signals/min. Interval between signal and nearest response = .25 sec.

Ln	nF	
	−.5 Sec.	.5 + Sec.
−.5 Sec.	.42	.42
.5 + Sec.	.49	.44

Mean = .44 sec.
No. of cases = 623.

and the *nearest* event after it. Several examples are given in Fig. 2. In these examples, F and L, are respectively, the first and last events of the sequence, and n is the nearest event before F and after L. In every case the two measurements nF and Ln were made. To avoid overdivision of the data, the measurements were grouped into a fourfold classification, viz:

Interval nF: shorter than .5 sec.
longer than .5 sec.
Interval Ln: shorter than .5 sec.
longer than .5 sec.

The measurements were further classified according to whether or not the relevant signal in the sequence being considered was missed. A 2-min. sample from each *S* was examined.

Evidence in favor of the hypothesis would be considered to exist if it could be shown that for a given signal speed and a given sequence the proportion of signals missed increased when nF and/or Ln were short. If the proportion of signals missed was independent of the size of nF and Ln, then the hypothesis would be negated insofar as the criteria expressed by the terms nF and Ln were valid. Table 6 shows the proportion of signals that were missed in each of the four categories of temporal arrangement of events outside of the sequence.

In no case in Table 6 is the difference between any two proportions significant at the 5% level. Whether or not a signal is dealt with seems to be independent of the position of events outside the sequence. Thus it seems unlikely that the increased proportion of omissions that accompany speed increases can be *directly related* to the increased density of events in the series.

The procedure described above was repeated for the signal speed of 80 signals/min. and the same size of interval (.25 sec.). Table 7 gives the results of the analysis of 315 cases.

No proportion in Table 7 is significantly different from any other. Thus the data in Table 6 are confirmed for a different speed. But Table 7 serves another purpose. If the proportion in each cell in Table 7 is tested against the comparable value in Table 6, it is found in each case to be significantly smaller. This means that for a given sequence, when the temporal arrangement of events around a sequence is constant, a signal is missed proportionately more often at a faster than at a slower mean speed. Thus, although the temporal position of events beyond the sequence seems not to be

TABLE 7

PROPORTION OF SIGNALS MISSED CLASSIFIED ACCORDING TO THE TIME INTERVAL BETWEEN THE SEQUENCE AND THE NEAREST EVENTS

Signal speed: 80 signals/min. Interval between signal and nearest response = .25 sec.

Ln	nF	
	−.5 Sec.	.5 + Sec.
−.5 Sec.	.20	.27
.5 + Sec.	.20	.25

Mean = .22 sec.
No. of cases = 315.

important for behavior relative to that sequence, the mean signal speed is. This appears to suggest the presence of a form of stress, which accompanies increases in mean speed but which is not directly related to the increased density of events that also occur. This stress is called *speed stress*, and is considered to be a general environmental stress which has a strength that is roughly constant if the mean signal speed is constant, and is independent of detailed changes in the prevailing temporal conditions. On the one hand, evidence in support of a hypothesis that relates omissions to the temporal arrangement of events outside the sequence is lacking; on the other hand, evidence is presented that supports an alternative hypothesis to explain why proportionately more signals are missed as the mean speed is increased. It should be remembered, however, that speed increases are not effective only in this generalized fashion.

Table 5 shows that for a given size of interval between a signal and the nearest response proportionately more signals will be missed at faster mean speeds. But at faster mean speeds, more signals per unit time will be close to a response merely as a result of the distribution of the interval between signals.

Summary

To explain why speed increases in the demanded rate of work in a sensori-motor skill led to a disproportionate deterioration in performance, specific association was sought between signals omitted and definable temporal relationships occurring as a result of continuous display changes. The distribution of the time interval between the signals presented was such that some relationships occurred many more times than others, the frequency being determined by the mean signal speed.

A detailed examination of performance in these terms revealed that omissions occurred close to the nearest response more often than would be expected by chance, and they were as likely to occur just before the response as just after. Furthermore, the probability that a signal would be omitted when it occurred at a constant time interval from a response increased linearly with mean speed.

The effect of increasing the signal speed was twofold. First, it increased the chances that a signal would occur close to a response, thus subjecting it to a hazardous temporal situation. Second, if a signal did occur close to a response, it increased the chances that it would be omitted. This latter effect was not due to the crowding in of other specific events as the speed increased, but to the general effect of a shortage of time in the situation as a whole.

There are certain implications contained in these conclusions that would appear to be highly relevant particularly to the study of perception in skilled behavior. It seems likely that in continuously changing displays, the character of the signal that stimulates response and the character of the response may be equaled in theoretical importance by the temporal relationships that involve these events.

References

1. Bills, A. G. Blocking: a new principle of mental fatigue. *Amer. J. Psychol.*, 1931, **43**, 230.
2. Conrad, R. Speed and load stress in a sensori-motor skill. *Brit. J. industr. Med.*, 1951, **8**, 1–7.
3. Conrad, R. The accuracy of response in a sensori-motor skill. *Med. Res. Coun., Appl. Psychol. Unit.*, 1953, Rep. No. 186.
4. Cox, D. R., & Smith, W. L. The superposition of several strictly periodic sequences of events. *Biometrika*, 1953, **40**, 1–11.

(Received October 19, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

PSYCHOMOTOR RESPONSE ACQUISITION AND TRANSFER AS A FUNCTION OF CONTROL–INDICATOR RELATIONSHIPS [1]

JACK A. ADAMS

Skill Components Research Laboratory, Air Force Personnel and Training Research Center

Psychomotor tasks frequently have the following characteristics: (*a*) a visual display consisting of a stimulus and an indicator; (*b*) a control mechanism that the operator must manipulate in response to the stimulus. The indicator is linked to the controls and responds in some specified manner to control movement. Typically, a task requires the indicator to be kept in alignment with, or superimposed on, the stimulus. Failure to achieve this task goal is revealed by a discrepancy between the stimulus and the indicator, and the operator must then eliminate this discrepancy by appropriate control movement. Proficiency in executing such responses has been shown to be importantly related to the relationship between control movement and consequent indicator movement (**1, 3, 5, 6, 7, 8**).

The investigation to be reported had a twofold interest in control-indicator relationships: (*a*) the effect of control-indicator relationship on the acquisition of a psychomotor response, and (*b*) the transfer of response acquired on a task with a specified control-indicator relationship to a variation of the same task wherein the control-indicator relationship had been changed. In this latter aspect of the experiment, the changes were all in the direction of indicator movement, with the direction of control movement remaining the same. The effect of indicator change on the amount of motor-response transfer represents a somewhat different task variable than those considered by Lewis, McAllister, and Adams (**4**) and Duncan (**2**) in their studies of transfer. Lewis *et al.* evaluated the effect of change in the direction of control movement, and Duncan varied the number of S–R connections from the first to the second task in a paired-associates learning situation. Both of these investigators were concerned with variation on the response side of the task (display unchanged), whereas the research to be reported deals with variation on the display side (response unchanged).

Method

Subjects.—The *S*s were 142 basic airmen trainees selected from the population available at Lackland Air Force Base, Texas. They were tested individually and were randomly assigned to five treatment groups.

Apparatus.—The Pedal Control Test was used. The *S* was seated, and he operated a pedal with each foot. On a panel in front of him and at about eye level were two 3½ × 7-in. rectangular display boxes containing the stimulus lights and the indicator lights. Angular orientation of these display boxes could be varied in 45° steps. Each box had a row of ten red stimulus lights and a parallel row of ten green indicator lights. The row of green lights in a box came on in succession as its corresponding pedal was depressed. The green lights in the left box were operated by movement of the left pedal, and the green lights in the right box were operated by the right pedal. A single red light would come on in each of the boxes, and the task of *S* was to match each red light with a green

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Air Force.

light by appropriate movements of the pedals. When this match had been achieved and held for .5 sec., the pair of red lights was immediately withdrawn and replaced by a new pair, and *S* repeated the matching procedure. Twenty-five random pairs of red lights were used, and this order of 25 was continuously repeated. The score was the number of matches made in a trial period.

Procedure.—To evaluate response acquisition, a control group (C) and four experimental groups (I–IV) were administered 15 trials on which each group had a different angular orientation of the two display boxes (Task A). Thus, each group practiced with the green indicator lights "moving" in a different direction with respect to the direction of control movement. Direction of control movement was the same for all groups, and the two display boxes were always parallel regardless of angular position.

Following Task A, Groups I–IV were given 15 additional trials on Task B where the control-indicator relationship was the same as for Group C in Task A. The comparison of performance curves for Groups I–IV on Task B with Group C's curve on Task A permits an evaluation of transfer as a function of control-indicator relationship in prior training.

To schematize the specific control-indicator relationships employed for each group in Task A, let the top of the display boxes be designated as 0° when they are in the vertical position. For Group C, movement of the green indicator lights was in the same direction as the pedal movement. Thus, when the pedals were depressed, the green lights moved from 0° toward 180°, and when the pedals were released, the spring tension on the pedals returned them toward 0°. Group I had the boxes turned 45° to the left such that depression of the pedals made the green lights move toward 135°. Group II had the boxes turned 45° further to the left than Group I such that the two boxes were horizontal and the green lights moved toward 90° when the pedals were pressed down. For Group III the green lights moved toward 45° when the pedals were depressed, and for Group IV green light movement was toward 0°. It can be seen that for this latter group, control movement and movement of the green indicator lights were in opposite directions.

Trial periods were 2 min. in length with 15-sec. rest periods between trials. For Groups I–IV this intertrial rest period was increased to 1 min. between Tasks A and B to permit changing of display box orientation. The *N* for each group was as follows: C: 30, I: 29, II: 27, III: 30, IV: 26.

Results

The performance curves for the five groups on Task A are shown in Fig. 1, where the performance measure is number of matches made in a trial period. Each group practiced with a different control-indicator relationship, and Fig. 1 shows little difference between groups in trend or average level. A test of these differences was made by trend analysis, and a summary table is presented in Table 1. The pooled between-*Ss* mean square was used as the error term to evaluate groups mean square as a test of the null hypothesis for mean performance level over these trials. The value of *F* for this test was less than one and did not permit a rejection of the null hypothesis. The hypothesis of parallel trend for the five curves was tested by using the pooled Trials × *Ss* mean square as the error term for evaluating the Trials × Groups mean square. The obtained *F* lacked significance at the 5% level of confidence, with a consequent retention of the hypothesis of same trend.

Transfer as a function of control-indicator relationship in prior training is shown in Fig. 2. The performance curves for Groups I–IV on Task B are shown where practice was with Group

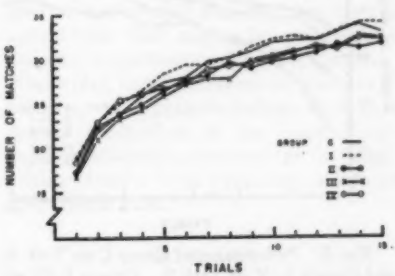


FIG. 1. Performance of control group (C) and four experimental groups (I–IV) on Task A, where each group practiced with a different control-indicator relationship

TABLE 1

TREND ANALYSIS OF PERFORMANCE ON THE 15 TRIALS GIVEN ALL GROUPS ON TASK A

Source	*df*	Mean Square	*F*	*p*
Trials	14	2,751.07		
Groups	4	236.93	.71	>.05
Trials×Groups	56	7.99	1.29	>.05
Pooled Between−*Ss*	137	332.61		
Pooled Trials×*Ss*	1,918	6.21		
Total	2,129			

C's control-indicator relationship. Group C on Task A is the control condition, and its performance curve is also presented in Fig. 1. Initial trials show a positive transfer effect for all experimental groups since they have a higher performance level than Group C. A simple analysis of variance of the means on Trial 1 gave an *F* of 36.80, which for 4 and 137 *df* was significant beyond the 1% level of confidence.

It is interesting to note that the amount of initial positive transfer is directly related to the correspondence between the control-indicator relationships of Task A and Task B. Group I had the least difference in control-indicator relationships from Task A to Task B, and Fig. 2 shows that it has the greatest amount of initial positive transfer. Groups II, III, and IV have successively larger differences in control-indicator relationships from Task A to Task B, and they have successively smaller amounts of initial positive transfer. However, systematic relationship between amount of positive transfer, and correspondence of control-indicator relationship between the two tasks, breaks down with training. Groups I, II, III, and C all appear to converge to a common level. A test of the null hypothesis for the means of these four groups on Trial 15 was made by simple analysis of variance and gave an *F* of 1.56, which for 3 and 112 *df* lacked significance at the 5% level. Group IV shows a trend decidedly different from the other groups. Although Group IV exhibits initial positive transfer, its curve crosses Group C's and appears to be stabilizing at a level lower than that of the other groups. When Group IV is included with the other four groups in a test of the null hypothesis for means on Trial 15 by simple analysis of variance, the obtained *F* of 2.94 for 4 and 137 *df* is significant at the 2% level.

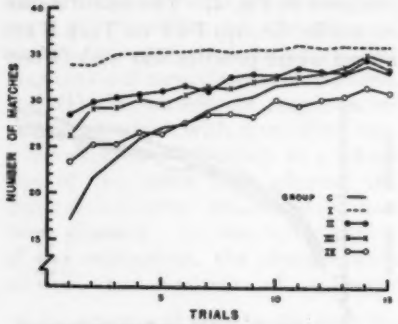


FIG. 2. Performance of Group C on Task A, and Groups I–IV on Task B. Groups I–IV are practicing with Group C's control-indicator relationship, but each has had prior training on the task with a different control-indicator relationship.

To give an over-all statistical evaluation of the five curves in Fig. 2, a

TABLE 2

TREND ANALYSIS OF PERFORMANCE ON THE 15 TRIALS GIVEN GROUP C ON TASK A AND GROUPS I–IV ON TASK B

Source	*df*	Mean Square	*F*	*p*
Trials	14	858.23		
Groups	4	3,589.41	9.32	<.01
Trials×Groups	56	71.84	12.56	<.01
Pooled Between−*Ss*	137	385.28		
Pooled Trials×*Ss*	1,918	5.72		
Total	2,129			

trend analysis was made. The results of this analysis are given in Table 2. The same hypotheses were tested as in the trend analysis presented in Table 1. Table 2 shows that the null hypothesis, both for between-groups differences and for trend, can be rejected beyond the 1% level of confidence.

Discussion

The absence of differences in trend and mean level for the acquisition curves suggests that no preferences (population stereotypes) exist for the control-indicator relationships investigated in this study. This finding is noteworthy in view of consistent preferences found when hand controls are employed (**1, 3, 5, 6, 7, 8**), and it suggests the hypothesis that population stereotypes are found only for a bodily member used for response in such control-indicator systems as are commonly found in everyday life. For example, hand movement of a wheel to the right to move an indicator to the right is a system encountered in automobiles, lathes, etc., and a preference for such control-indicator relationships might be expected on other devices using a similar system. In the case of pedal control-indicator systems, few are in common use and it would seem plausible to assume that little preference exists in the population with respect to them. Foot responses are more generally involved in kinesthetic discriminations such as those required for clutch or brake operation.

With respect to transfer of training, these data indicate that when control-indicator relationships differ between two forms of a task, practice on one will result in initial positive transfer to the other, and the amount is directly related to the correspondence between the two control-indicator relationships. This initial advantage decreases with training and finally is reduced to no advantage (complete reversal of control-indicator relationship from one form of task to the other excluded). Thus, initial practice on a first task leads to more rapid attainment of some pre-asymptote performance level on a second task with a different control-indicator relationship, but does not affect final asymptote performance level.

Although the performance data for Group IV are interesting in showing the over-all effect of reversing the control-indicator relationship, they are inadequate for specifying the particular response characteristic most affected by this change. Response acquisition on the Pedal Control Test not only involves learning the direction of movement, which is a two-alternative decision, but also the learning of the appropriate response amplitude, which can assume step values of one to nine from the previous light. On a priori grounds it might seem that most of the response variability could be ascribed to the learning of response amplitude, with relatively little related to the learning of response direction. As a consequence, negative transfer would mainly tend to be reflected in amplitude errors, with a lesser or negligible effect on initial choice of direction. However, these response measurements were not taken, and such determinations are therefore not possible within the framework of this experiment. Future research must answer these more detailed questions on negative transfer.

Studies of transfer bear importantly on problems in industrial and military training programs where questions arise on the effect of differences between a training task and an operational task on amount of transfer. In order that the goal of minimizing training time on the operational task might be achieved, it becomes important to delineate the effect of specified task differences on amount of transfer. The present findings show that careful simulation of the control-indicator relationship is important for maximizing transfer from a training task to an operational task.

Summary

The Pedal Control Test was used to evaluate the effect of control-indicator relationship on psychomotor response acquisition and on transfer of training to another form of the same task

where the control-indicator relationship had been changed.

Response acquisition was found to be independent of the five control-indicator relationships investigated. This suggested the hypothesis that bodily members that are infrequently used for response in control-indicator systems found in everyday life do not show preference for control-indicator relationships.

With respect to transfer of training, it was found that when control-indicator relationship differed between two forms of the task, practice on one resulted in initial positive transfer to the other, and the amount was directly related to the correspondence between the two control-indicator relationships. Training on the second task decreased this initial advantage until, at the end of training, three of the four experimental groups were not significantly different from the control group. A fourth experimental group, which had complete reversal of control-indicator relationship from one form of the task to the other, exhibited negative transfer in the later stages of training.

References

1. Carter, L. F., & Murray, N. L. A study of the most effective relationships between selected control and indicator movements. In P. M. Fitts (Ed.), Psychological research on equipment design. Washington, D. C.: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 19.) Pp. 147–157.
2. Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1–11.
3. Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1287–1340.
4. Lewis, D., McAllister, D. E., & Adams, J. A. Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–260.
5. Mitchell, M. J. H. Direction of movement of machine controls: III. A two-handed task in a discontinuous operation. Unpublished report, Applied Psychology Unit, Cambridge University, 1947.
6. Mitchell, M. J. H. Direction of movement of machine controls: V. A two-handed performance in a continuous task. Unpublished report, Applied Psychology Unit, Cambridge University, 1949.
7. Mitchell, M. J. H., & Vince, M. A. The direction of machine controls. Unpublished report, Applied Psychology Unit, Cambridge University, 1951.
8. Warrick, M. J. Direction of movement in the use of control knobs to position visual indicators. In P. M. Fitts (Ed.), Psychological research on equipment design. Washington, D. C.: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 19.) Pp. 137–146.

(Received October 5, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

MULTIPLE VERSUS SINGLE PROBLEM TRAINING IN HUMAN PROBLEM SOLVING[1]

JACK A. ADAMS

Skill Components Research Laboratory, AF Personnel and Training Research Center

Military training programs for maintenance technicians must be concerned with methods of training proficient problem solvers who can quickly trouble-shoot malfunctioning equipment. Harlow (1) has presented evidence to show that highly effective problem-solving behavior can be acquired through practice on a large number of problems where all problems have the same general solution but differ in stimulus properties from problem to problem. He has termed the learning of the ability to readily solve a new problem of a specified class as the formation of a "learning set."

The study to be presented investigates the effectiveness of two different training procedures in the formation of a learning set. Harlow's method of training on a large number of problems may be only one of several possible training techniques that might be used. One possibility is that repeated presentation of the *same* problem could also lead to considerable facility in the solution of a new problem. This investigation evaluates the relative effectiveness of multiple problem and single problem training in the solution of a simple discrimination problem by adult human *S*s.

METHOD

Apparatus.—Stimuli were on 2 × 2-in. slides and presented by a Selectroslide projector manufactured by Spindler and Sauppe of Los Angeles, California. The *S*s were seated between the projector and the screen and were run four at a time. The table at which *S*s were seated was partitioned so that they could not see one another respond. In front of each *S* was a small response box on which were mounted four horizontal and linearly arranged plunger-type microswitches and a small frosted lamp. When a slide was presented, *S* pressed the one of the four buttons he judged correct. If the choice was correct, the lamp on the response box came on.

Subjects.—The *S*s were 127 basic airmen trainees selected from the population available at Lackland Air Force Base, Texas. Subgroups of four *S*s were randomly assigned to a multiple problem training group (M) and a single problem training group (S). The *N* was 68 in Group M and 59 in Group S.

Experimental procedure.—The basic unit of study was the *problem*, which was comprised of eight slides. These eight slides consisted of two slides each of four different spatial arrangements of a stimulus pair, and the particular arrangement of any given pair indicated which of the four buttons on *S*'s response box was correct. For example, consider a problem involving two stimulus forms, A and B. The solution to this problem is: A to the left of B, Button No. 1; A to the right of B, Button No. 2; A above B, Button No. 3; and A below B, Button No. 4 (buttons numbered from left to right). These relationships were maintained throughout all problems regardless of the stimulus pair used. A sophisticated *S* who "understands" these relationships could conceivably respond 100% correct following the first slide of a new problem with Stimuli C and D. If the new pair is arranged in a horizontal pattern on the first slide, then *S* "knows" that either Button No. 1 or 2 is correct. If he chooses Button No. 1 and it is incorrect, then he knows that No. 2 must be correct, and correct responses for the other three pairs to be presented on the subsequent seven slides of the problem can be immediately deduced. Similarly, if the stimulus pairs are arranged vertically on the first slide, *S* knows that either Button No. 3 or 4 is correct, and the correct response for each of the remaining slides of the problem can be deduced from his choice on the first slide. This sophistication

[1] The experimental work for this study was performed as part of the United States Air Force Personnel and Training Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

in the solution of this simple problem is, however, the type of behavior to be acquired, and the training method for maximizing it is the factor under investigation in this study.

Each *S* was presented 216 slides. There were 192 slides in the training series and 24 in the test series. Group M received 24 *different* problems in the training series and three presentations of the twenty-fifth problem for the test series. Group S received 24 presentations of the *same* problem in the training series and then received the same test series as Group M. The stimulus pair in the problem given Group S in the training series was a circle and a triangle. Problem 1 for Group M was also a circle and a triangle so that the equivalence of groups might be determined prior to differential treatment. Problems 2–24 for Group M involved abstract stimuli. The test problem given both groups also had abstract stimuli. Care was taken to minimize stimulus similarity within a problem and between problems. The eight slides were randomized separately for each problem with the restriction that no two identical slides follow one another.

Instructions were quite general and indicated only that a series of stimulus pairs would be presented and for each pair *S* had to learn to turn on the lamp on the response box by selecting the correct button. These initial instructions were the only instructions given and were the same for both groups.

A slide remained on the screen until all four *S*s had responded. The *S*s were under instruction to respond only once to each slide and to respond rapidly. Under this speed set, *S*s typically responded within 3 sec. after the slide first appeared on the screen. In order that all *S*s might have time to observe the lamp on the box and note correctness of their responses, a slide was not removed until 3 sec. after the last *S* had responded.

It could be contended that the speed set employed would not allow *S*s sufficient time to fully consider all possible hypotheses before responding, with the result that maximum proficiency in problem solving may not be exhibited. Although there may be some merit in this contention, the purpose of this experiment was to compare two methods for training in problem solving, and there is no reason to believe that the speed set interacted with the variables studied.

A 2-min. rest was given between each block of six problems. All training was completed in a single practice session, and both groups took about the same amount of time to complete their respective slide series.

Results

Figure 1 shows the percentage correct for each problem given Groups M and S. Problem 1 was the same for both groups and both have 27% correct. With chance response level at 25% (four possible responses), the obtained 27% does not exceed chance at the 5% level of confidence. Both groups steadily increase in proficiency over the training series. Group S, which was trained on repeated presentation of the same problem, is superior to Group M and shows a negatively accelerated acquisition curve. Group M, which was given multiple problem training, has an essentially linear increase in percentage correct as a function of number of training problems. It should be noted that the limit of the acquisition curves in the training series differs for the two groups. The limit is 100% for Group S but only 94% for Group M because at the asymptote *S*s in Group M would average 50% correct on the first slide of a new problem.

In the test series, both groups were given three presentations of a new problem to evaluate their ability to solve it as a function of type of prior training. Figure 1 shows that both groups have the same value of 61% correct on the first presentation. For Group M this test problem is only

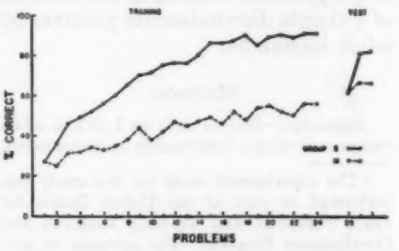


Fig. 1. Performance of the multiple problem training group (M) and the single problem training group (S) on the training and test problems

another new problem, and 61% correct is consistent with the trend established in the training series. However, for Group S the test problem is the first new problem encountered and their level of 61% on the first presentation is a 30% reduction in proficiency from the level on the final training problem. On the second and third presentations of the test problem, Group S recovers most of this loss and is clearly superior to Group M. A test of the null hypothesis for the difference between the two groups gives a probability of .06 for the second presentation and .02 for the third presentation.

To examine test series performance more closely, Fig. 2 represents a detailed enlargement of the three test series values shown in Fig. 1 and gives the percentage correct for each group on each of the 24 slides presented in the test series. The important feature to note in this graph is that Group S undergoes a brief period of adjustment to the new problem during which performance level is below that of Group M. On Slide 1, Group S's level of 34% is not significantly above the chance level of 25% ($p = .12$). A test of the null hypothesis for the difference between the two groups on Slide 1 yields significance at the 6% level of confidence. On Slide 2, the difference between the two groups is significant at the 2% level. By Slide 3 there is negligible difference between them and, on subsequent slides, Group S displays consistent superiority.[2]

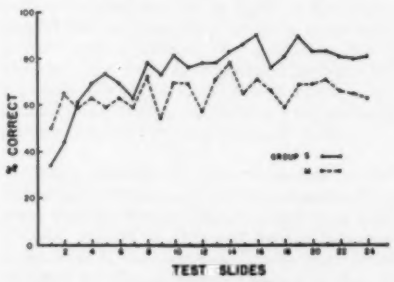


FIG. 2. Performance of Groups M and S on each slide of the test series

The data for both groups were examined for individual differences in problem-solving ability. Group S was divided into low- and high-ability groups of 17 *Ss* each; these were compared for their ability to solve the new problem in the test series. The high- and low-ability designation was on the basis of number of presentations of the training problem required before solution was obtained. The high-ability group solved the problem in 2–7 presentations and the low-ability group required 14–23 presentations. On presentations 1, 2, and 3 of the test problem the high-ability group had 57%, 85%, and 84% correct, and the low-ability group had 60%, 73%, and 79% correct. It is evident that ability differences in single problem solution are not importantly reflected in the solution of a new problem of the same class. Group M revealed no important information with respect to ability differences. The *Ss* who acquired solution of the problem class in the training series also readily solved the test problem since it was merely another problem of the same class. Those *Ss* who failed to achieve com-

[2] It could be contended that performance of Group S on the test problem was influenced by the sudden change from conventional symbols used in the training series (circle and triangle) to abstract forms used as test problem stimuli. To evaluate this possibility, Group S was replicated ($N = 56$), with the exception that abstract forms were used as stimuli in *both* training and test series. Performance trends of Group S and the replicate group were found to be essentially the same. A test of the null hypothesis was made for the difference between these two groups on each of the three presentations of the test problem. None of the obtained differences was significant at the 5% level.

plete solution at the end of the training series also failed in the test series.

Discussion

The general superiority of Group S in the test series indicates that, with the amount of training given, single problem training leads to greater over-all proficiency in the solution of a new problem than multiple problem training. It is noteworthy, however, that Group S is below the performance level of Group M on the first two slides of the test problem. This suggests that one advantage of multiple problem training is smooth transition to a new problem.

Group M's increasing ability to solve new problems is consistent with previous research (1) where multiple problem training resulted in considerable facility in the solving of new problems. The fact that Group M is only slightly above 50% by the end of training shows that an intermediate stage in the solution to this class of problems has been attained where Buttons 1 and 2 are associated with horizontally arranged stimuli, and Buttons 3 and 4 with vertically arranged stimuli. The stage of learning has not yet been reached where the correct response on all subsequent slides of a problem can be deduced from the response on the first slide. The linear increasing trend suggests that this high proficiency may be attained after a sufficient number of training problems.

Summary

Under investigation were two methods of training *Ss* in the solution of problems of a specified class. The problem class under consideration was a simple discrimination where the general solution was the same for all problems but the stimulus characteristics could vary from problem to problem. It was found that a group trained on repeated presentations of the same problem was more proficient in solving a new problem of the class than a group trained on a number of different problems.

References

1. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

(Received November 6, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

INCENTIVE CONDITIONS WHICH INFLUENCE VISUAL EXPLORATION [1]

ROBERT A. BUTLER

Walter Reed Army Medical Center

Discrimination learning by monkeys for no reward other than that provided by visual exploration of the environment was demonstrated by Butler (2) in a previous experiment. In this prior study, precautions were taken to prevent *S*s from seeing food or other monkeys during testing, but no attempt was made to control the nature or number of other events.

The present experiments were designed to measure the strength of visual exploratory behavior of monkeys under varied conditions of external stimulation. Experiment I investigates changes in performance associated with exposure to various categories of visual-auditory stimulation. Experiment II is essentially a repetition of the first experiment except that on some conditions the source of the auditory stimuli was removed from the prospective visual field of the animals.

METHOD

Subjects

Eight monkeys served as *S*s in Exp. I, and ten monkeys in Exp. II. These preadolescent animals weighed from 4½ to 7½ lb. at the beginning of the researches and had been tamed and adapted prior to testing. The animals in Exp. I had been run on a discrimination problem for approximately 60 days, and the animals in Exp. II had completed a fixed-ratio reinforcement problem requiring approximately 30 days.

Apparatus

The *S*s were housed during testing in a wire cage, measuring 27 × 17 × 26 in., which was covered by a box whose front and top were Masonite and whose three remaining sides were heavy black cloth supported on wooden frames. The box was raised and lowered by means of a pulley system fastened to the ceiling of the test room. A hinged door measuring 3¼ × 4 in. was centered in the face of the box (see Fig. 1), and a 3-in.-sq. transparent Lucite plate was attached to the inside face of the door with sufficient space to permit insertion of a white card. An opaque screen, which could be raised or lowered by *E*, was placed immediately behind the box face and door; this allowed *E* to permit or prevent door contacts as desired. Illumination for the cage, door, and card was provided by a 25-w. lamp attached to the roof of the box. Connected to and facing the front of the monkey cage was an incentive chamber measuring 48 × 36 × 30 in. A laboratory table top formed the floor of the chamber, top and sides were constructed by heavy black cloth on a wooden frame, and illumination was provided by a 100-w. lamp centered in the ceiling.

Temperature in the test room during Exp. I ranged from 60° to 82° F. and averaged 70.7° F. It ranged from 59° to 88° F., with a mean of 74.7° F., while Exp. II was in progress. Temperature inside the apparatus increased about 4° F. above room temperature during a 30-min. test session.

The incentives measured in Exp. I were: (*a*) another monkey confined in a transport cage (monkey condition); (*b*) a continually moving toy electric train consisting of an engine and two cars (train condition); (*c*) an array of food including a fresh apple, orange, banana, carrot, and several pieces of laboratory chow (food condition); (*d*) the empty incentive chamber (standard condition).

The incentives used in Exp. II were the same as in Exp. I for monkey, train, and standard conditions. Two additional conditions were employed: a hear-train condition, with the continuously moving electric train on the floor beneath the incentive chamber; and a hear-monkey condition, with the monkey in the transport cage on the floor beneath the empty incentive chamber and another monkey caged in an adjacent

[1] These experiments were conducted at the University of Wisconsin and were supported by a grant from the Atomic Energy Commission (AT 11–1–64, project no. 11) and by the Research Committee of the Graduate School from funds supplied by the Wisconsin Alumni Research Foundation, both under the general supervision of Dr. Harry F. Harlow.

FIG. 1. Visual exploratory response to train condition

room. To insure adequate auditory stimulation, the two most vocal monkeys in the laboratory were used. The hear-train and hear-monkey conditions consisted of a combination of the visual stimuli in the standard condition and the auditory stimuli from either the train or the monkeys.

It is to be emphasized that tests made with and without the stimulus source being present in the incentive chamber are not problems of cue additivity because the auditory cues in the monkey and hear-monkey conditions are not comparable.

Procedure

Preliminary training.—Preliminary training was initiated by removing the incentive chamber from the front of the test apparatus, placing *S* in the wire cage, and lowering the box over the cage. The *E* next raised the opaque screen and opened the door, permitting *S* to explore the test room visually. Periodically *E* partially closed the door. When *S* responded by reopening the door, *E* returned to the apparatus and closed the door farther.

After *S* repeatedly pushed against the completely closed door and opened it, formal adaptation trials were given. The *E* initiated a trial by raising the opaque screen and exposing the door to the monkey. If *S* opened the door, it was allowed 60 sec. to explore the test room visually before the opaque screen was lowered. If *S* did not respond to the door within 5 min. after the opaque screen was raised, *E* opened the door from the outside and waited 60 sec. before lowering the screen. Intertrial interval, defined as time between lowering and raising the opaque screen, was 30 sec. Preliminary training was continued until *S* made at least five responses to the door within a 30-min. test period for two successive days.

Test trial procedure.—The test trial procedure for Exp. I and II involved placing *S* in the cage with the incentive chamber attached, and then allowing 3 min. for adaptation. The *E* started each trial by raising the opaque screen. The monkey was permitted a maximum of 5 min. to open the door, and 5 sec. to explore the incentive chamber visually after the door was opened. The opaque screen was then lowered by *E*, and the door closed. If *S* failed to respond within 5 min., the door was opened from the outside by *E*, who waited 5 sec. before closing it. The intertrial interval, again defined as time between lowering and raising the opaque screen, was 10 sec. regardless of *S*'s response, or failure of response, to the door.

Experimental design.—The *S*s in Exp. I were tested 30 min. a day for five successive days on each of the four stimulus conditions described as monkey, train, food, and standard. A latin-square design was utilized, with four blocks of five test days assigned to columns, stimulus conditions to letters, and monkeys to rows of two replicated 4×4 latin squares. The dependent variable was total number of responses per block. Daily observational records were also made of cage shaking, vocalizations, defecation, and urination. A unit score was given for the presence of each of these conditions and a zero score for its absence. A composite measure was obtained from the sum of these scores for each monkey under each stimulus condition and was described as a general behavior score. The maximum gen-

TABLE 1

RESPONSE FREQUENCY AND GENERAL BEHAVIOR SCORES FOR THE DIFFERENT STIMULUS CONDITIONS

Condition	Response Frequency		General Behavior Scores	
	Mean	*SD*	Mean	*SD*
Experiment I				
Monkey	189*	66.6	10.5*	3.3
Train	128*	23.5	9.4*	4.5
Food	97	64.0	6.2	3.6
Standard	70	20.8	6.2	3.5
Experiment II				
Monkey	242	98.5	10.7	3.9
Train	213	97.9	11.3	4.6
Hear-Train	203	101.1	8.3	5.8
Hear-Monkey	159	81.2	12.7	4.8
Standard	161	101.3	9.7	5.6

* $p < .01$.

eral behavior score obtainable by a monkey under any one stimulus condition was 20, indicating that cage shaking, vocalization, defecation, and urination were present on each of the 5 test days.

Experiment II involved comparisons among the five stimulus conditions of monkey, train, hear-monkey, hear-train, and standard. The experimental design was identical with that employed in Exp. I, except for the use of 5×5 rather than 4×4 latin squares.

Results

The results for both Exp. I and II are presented in terms of frequency of responses and the general behavior scores (see Table 1). Data on temperature effects are not presented because no consistent relationship was found between temperature and performance from day to day.

Frequency of Response

The data of Exp. I were analyzed by Friedman's chi-square test for the differences between correlated means (**4**). Differences in response frequency obtained among the four stimulus conditions are significant beyond the .01 confidence level. The only significant differences ($p < .01$) between specific conditions are those between the monkey and standard conditions and those between the train and standard conditions. No consistent or significant change was found from a day-to-day analysis of response frequencies. The rank-order correlation between consecutive test days and the total number of responses for the respective days was .22 ($p > .05$).

The most striking finding in Exp. II was the day-to-day increase in response frequencies as the experiment progressed. Friedman's chi-square test indicated that the number of responses differs significantly ($p < .01$) from Day 1 to Day 25. The rank-order correlation between response frequency and days was .80 ($p. < .01$).

The significant increase in number of responses as a function of days seemed to suppress the influence of stimulus conditions on response frequency. The effect of days was removed, and the differences between stimulus conditions were retested for significance. Cancellation of the effect of days introduced certain statistical difficulties, which were resolved by using a cross-validation type of significance test. The ten *S*s were split into two groups (Groups A and B). One *S* in each group had experienced the same order of presentation of stimulus conditions. The daily means were calculated for Group A and were then subtracted from the corresponding scores of *S*s in Group B. An optimal linear function (see Rao, **5**, pp. 239–246), calculated on Group A, was applied to Group B. The mean of this function was significantly greater than zero ($p < .01$), indicating that the stimulus condition means were significantly different after the scores had been corrected for the effect of days.[2] No tests for significance between the standard condition and each of the other stimulus conditions were conducted.

General Behavior Scores

The general behavior scores for the various stimulus conditions of Exp. I decreased in magnitude in the same order as the response frequency scores. Differences in scores among the stimulus conditions, as determined by the Friedman chi-square test, are significant beyond the .01 confidence level. The standard condition differs significantly from both the monkey condition and the train condition at the .01 level. No significant differences be-

[2] The author wishes to acknowledge the valuable advice and assistance of Dr. Ardie Lubin, Army Medical Service Graduate School, Walter Reed Army Medical Center, on this statistical procedure.

tween the general behavior scores were found for those stimulus conditions presented in Exp. II.

Rank-order correlations between consecutive test days and the total general behavior score recorded for the respective days were $-.14(p>.05)$ for Exp. I, and $-.40(p>.05)$ for Exp. II.

Discussion

The data of Exp. I and II are consistent in demonstrating that visual exploratory responses persist for at least 20 and 25 days, respectively. These data are inconsistent in that response frequency of *Ss* in Exp. II increased significantly during the course of the experiment, whereas no consistent increase or decrease in frequency of responses was found in the first experiment. Differences in performance between *Ss* participating in Exp. I and those used in Exp. II might be attributed to the fact that the Exp. II animals had had less previous experience in test situations. That testing experience with the concomitant handling of the monkeys is an important factor for the monkeys' adapting to new test conditions is a general finding at the Wisconsin Primate Laboratory. Unlike the adult, test-wise monkeys used in previous visual exploration researches, most of the *Ss* employed in the present studies required several adaptation sessions before they would even approach the opened door attached to the front of the apparatus. Rather than look out of the door, the animals remained at the rear corner of the cage. It is suggested that *Ss* in Exp. II were subjected to formal test sessions before they had adapted adequately to the test situation, and that the increase in response frequency recorded for them was directly related to a decrease in the general apprehension associated with the testing procedures.

In Exp. I, the insignificant increment in responsiveness to food over the standard condition fails to support any argument that visual exploration is a derived hunger-food drive. The significant increment in responsiveness of monkey over standard condition leaves the operative factors indeterminate. There is a high probability that the operation of learned affections is present, but their exact nature and origin remain unknown. The high frequency of responses to the train condition obviously cannot be explained in terms of secondary motivation conditioned to a primary drive.

After correcting for the effect of days, the data of Exp. II agree with those of Exp. I insofar as significant differences in response frequency appear to be associated with differences in stimulus conditions. The results of the second experiment indicate that auditory stimuli, whose source is visually inaccessible, elicit visual exploratory responses, but leave quantitative comparisons between the different stimulus conditions indeterminate.

These data could be explained in terms of anxiety reduction. For example, if enclosing a monkey in the test apparatus produces anxiety, the opening of the door might serve to reduce anxiety, which, in turn, would reinforce the door-pushing response. Brown's interpretation of the results obtained in previous visual exploration studies implies an anxiety-reduction hypothesis to account for the monkeys' performances (1, p. 54). Records were kept on the general behavior of the monkeys for the purpose of finding out whether or not the factors thought to be related to anxiety were indeed associated with response frequency. The general behavior scores and the response frequency recorded for Exp. I followed the same trend with respect to the stimulus conditions. In Exp. II, however, the general behavior scores failed to increase concomitantly with frequency of responses as the experiment progressed. Actually, these scores were negatively correlated with consecutive test days. A more important factor relevant to an anxiety-reduction hypothesis was the tendency for monkeys to hide in the rear of the apparatus during the first several preliminary training sessions, which suggests that an anxiety state is antithetical to visual exploratory responses.

It is hypothesized that a visual exploration motive can best account for these data, and the persistence of responses obtained in this and other studies (3) suggests that visual exploration is as basic as any primary drive.

Summary

The influence of different external incentive conditions on visual exploration as measured by frequency of response was tested on 18 monkeys.

In Exp. I eight monkeys were tested on the four incentive conditions of viewing and hearing another (caged) monkey, viewing and hearing a moving electric train, viewing an array of foods, and viewing an empty incentive chamber. Response frequency was highest for the monkey conditions and decreased progressively for the other conditions in the order given.

In Exp. II ten monkeys were tested on the five conditions of viewing and hearing another monkey, viewing and hearing a moving electric train, and viewing an empty incentive chamber with and without the presence of sounds from a moving electric train or from other monkeys. Visual exploratory response rate continued to increase throughout the experiment. Significant differences among conditions were obtained after the effect of days was removed.

The data are interpreted in terms of a visual exploration motive whose persistence suggests that visual exploration is not conditioned upon, or derived from, any other primary drive.

References

1. Brown, J. S. Comments on Professor Harlow's paper. In J. S. Brown, H. F. Harlow, O. H. Mowrer, T. H. Newcomb, V. Nowlis, & S. L. J. Postman, *Current theory and research in motivation*. Lincoln: Univ. of Nebraska Press, 1953.
2. Butler, R. A. Discrimination learning by rhesus monkeys to visual-exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46**, 95–98.
3. Butler, R. A., & Harlow, H. F. Persistence of visual exploration in monkeys. *J. comp. physiol. Psychol.*, 1954, **47**, 260–265.
4. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
5. Rao, C. R. *Advanced statistical methods in biometric research*. New York: Wiley, 1952.

(Received September 21, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

LEARNING VARIABLES IN THE JUDGMENT OF SINGLE STIMULI [1]

ALLEN PARDUCCI

University of California

There is ample evidence that the so-called "absolute" judgment of a stimulus is not absolute, but rather is dependent upon various characteristics of the sequence of stimuli to which the judged stimulus belongs. Helson (1, 2) isolates three factors which, along with the physical value of the judged stimulus, determine its judgmental response: (*a*) the physical values of the other stimuli in the series being judged, (*b*) background stimuli present at the time of judgment, and (*c*) *S*'s past experience with similar stimuli. Additional variables, not yet included in Helson's formulae, are investigated in the experiments to be reported in this paper.

The selection of the independent variables was made in accordance with the general hypothesis that judgments are made in relation to what *S* has learned about the distribution of stimuli presented for judgment. If, for example, he has learned that these stimuli fall within a circumscribed range, stimuli outside this range will be assigned the extreme values of his judgment scale. By manipulation of *S*'s information about the distribution of stimuli, systematic changes in judgment should be produced; that is, variables which influence the nature and amount of information about the stimuli should also influence the distribution of judgments.

Experiment I

The first experiment tested the hypothesis that prior verbal description of the distribution of the stimuli would minimize shifts in judgment during their subsequent presentation.

Method

Subjects.—The *S*s were 25 college students in elementary psychology, 12 in the prior-description group and 13 in the control group.

Materials.—The stimuli were ten 5 × 8-in. cards, with 3-in. vertical lines, one red and one blue, on each card. These lines were spaced from ½ to 5 in. apart in ½-in. steps, with the blue line in a fixed position 1½ in. from the right edge of each card.

Procedure.—Each group (the two groups were run at different hours) was given response sheets and was read the following instructions:

"This is an experiment in judging the size of different stimuli. You will first be presented with a series of cards like this (*E* holds up the first card), each containing a red and blue line. As each card is presented, you make a judgment of how large or small the separation between the lines appears to you. Record your response in terms of one of the five size categories listed at the top of the page, using the abbreviations shown. Please respond to every card even if the decision seems difficult with the first ones."

The following description was then read (twice) to the experimental group only:

"The blue line will always appear here (*E* points to it), but the red line will be shifted from card to card. It will appear just as often in each of ten different positions. These ten positions are evenly spaced in ½-in. steps all the way across the card. The smallest separation will be ½ in.; and the largest, 5 in., will be when the red line is placed symmetrically with the blue one on the opposite side of the card."

Then, the following instructions were added for both groups:

"One thing more—do not go back and change your responses after new stimuli have been presented. This experiment concerns first impressions only; and since you are to judge how large the card looks to you, there can of course be no 'right' or 'wrong' answers. Be sure to use the abbreviations for the size categories printed at the top of the page."

After these instructions, the stimuli were displayed manually to *S*s, being held one at a time

[1] The writer is indebted to Dr. Leo Postman for many helpful suggestions concerning these experiments.

TABLE 1

EXP. I: ORDER OF PRESENTATION*

1–5 CBDEA	6–10 DCAEB	11–15 CBDEA	16–20 HGIJF	21–25 IHFJG	26–30 HGIJF
31–35 IHFJG	36–40 CHDFE	41–45 AGIBJ	46–50 CHDFA	51–55 EGIBJ	56–60 CBDEA

* Stimuli lettered in order of increasing size.

before a gray, 15 × 30-in. backdrop. There were 60 separate presentations, six of each card, in the order shown in Table 1. This was the same for both groups and was designed to create (for the control group) an initially false impression of the distribution of stimuli. Since the final presentations (55–60) are a repetition of the initial five stimuli, the hypothesis of this experiment predicts a greater shift in the judgments of these stimuli by the control than by the prior-description group.

Results

The judgments were converted to numerical form: very large = 5; large = 4; medium = 3; small = 2; very small = 1. Then the sum of each *S*'s last five judgments was subtracted from the sum of his first five judgments. The mean and variance of these differences were then calculated for both groups. The variance of the control group proved to be significantly larger than the variance of the prior-description group ($F = 4.2$; $p = .05$). While this was consistent with the hypothesis since it reflected the fact that there was very little change in judgment by any *S*s in the experimental group, it necessitated a logarithmic transformation of the scores in order to permit a t test for the difference in means. The means and *SD*'s based on these transformed scores are presented in Table 2. The transformed variances were almost equal ($F = 1.06$), and the difference in means supports the experimental hypothesis ($t = 2.27$; $p = .025$, one-tailed test). The possibility of finding a greater difference was limited by the fact that 5 of the 13 control *S*s showed no shift in judgment.

TABLE 2

EXP. I: MEAN SHIFT IN JUDGMENT*

Group	*N*	Mean	*SD*
Experimental	12	.480	.188
Control	13	.647	.182

* Log transformation of scores.

EXPERIMENT II

The second experiment tested the hypothesis that the anchoring effect of a stimulus would depend upon its nearness in the series to the judgment it was anchoring. This hypothesis is suggested by the principle of memorial *recency*—other things being equal, what is learned later (and has less chance of being interfered with) is remembered better than earlier learning. To test this hypothesis, a set of extreme or anchoring stimuli was introduced, without special instructions, either *early* or *late* in the series of regular stimuli. These anchors, according to the hypothesis, should have greater effect upon the final judgments of the regular stimuli when introduced in the later position. The direction of this anchoring effect would depend upon whether the extreme stimuli selected as anchors were above or below the range of the regular stimuli. Values at both high and low extremes were used. A subsidiary hypothesis, that the effect of serial

position would be less pronounced with more familiar materials, was tested by comparing this effect upon judgments of two different kinds of stimuli.

Method

Subjects.—Introductory psychology students, 128 in number, were assigned to eight groups of 16 *Ss*. Assignments were made according to the times *Ss* were free to serve.

Materials.—The stimuli for half the *Ss* were 11 square black cards, ranging in width from 7 to 17 cm. in 1-cm. steps. The remainder of the *Ss* judged the respective heights of sketches of men drawn on 11 × 15-in. sheets of white cardboard. Although all the sketches were approximately 10 in. high, they were made to appear different in height by varying the proportions between the size of the head and the rest of the body, the "taller" men having relatively smaller heads. The only other cue to height was the size of a set of steps which each man was depicted as descending; the "taller" men were placed on smaller steps. Seven of these men were drawn by trial and error so that each man appeared to be about 2 in. taller than the next shorter man.

Procedure.—The *Ss* were instructed to judge the stimuli either in terms of a five-category size scale or in terms of a corresponding height scale (depending upon whether they were judging the cards or the men) much as in Exp. I. The three smallest cards served as the low anchors and the three largest as the high anchors. The middle five cards, presented four times in different orders, made up the rest of the stimulus series for the four card-judging groups. Since only seven sketches of men were drawn, only the "tallest" and "shortest" could be used as anchors (and still leave five sketches for the regular stimulus series, as with the cards). To prevent perceptual isolation due to repetition of these anchors, the two end sketches of the regular series were also presented in the interpolated anchor series. This procedure can be represented by assigning a letter to each stimulus in the order of its size. Thus, in Table 3, the low anchors for cards were *a*, *b*, and *c*, and the high anchors were *i*, *j*, and *k*; the low anchors for men were *c* and *d*, the high anchors were *h* and *i*; and the regular series consisted of the middle five stimuli in each case (*d*, *e*, *f*, *g*, *h*). The order of presentation of the stimuli for the different groups is shown in Table 3.

TABLE 3

EXP. II: PRESENTATION OF STIMULI

Groups and Anchor	Order of Presentation				
	1–5	6–10	11–15	16–20	21–25
Cards					
Low-early	feghd	cbcba	gfdhe	dhgef	dhgef
High-early	feghd	ijijk	gfdhe	dhgef	dhgef
Low-late	feghd	gfdhe	dhgef	cbcba	dhgef
High-late	feghd	gfdhe	dhgef	ijijk	dhgef
Men					
Low-early	feghd	cdcdc	gfdhe	dhgef	dhgef
High-early	feghd	ihihi	gfdhe	dhgef	dhgef
Low-late	feghd	gfdhe	dhgef	cdcdc	dhgef
High-late	feghd	gfdhe	dhgef	ihihi	dhgef

Results

The *Ss*' judgments were transformed into numbers just as in Exp. I, and a comparison was made between the groups on the basis of *Ss*' mean judgments of the first five stimuli. It was found that neither the means of the four card-judging groups nor the means of the four men-judging groups differed significantly among themselves; thus the groups could be regarded as equated before the introduction of their respective anchors. It was also found that the variance between *Ss*' mean judgments of these five initial stimuli was significantly greater for the cards than for the sketches of men ($F = 1.68$, $p = .02$). This difference indicates that before any anchors were presented, *Ss* were in greater agreement about the height of the men than about the size of the cards; and it is consistent with the assumption that *Ss* would have greater familiarity with the former. That this difference in intersubject agreement had disappeared by the end of the experiment was indicated by application of Bartlett's test for homogeneity of variance to the mean judgments of the last five stimulus presentations by the eight groups ($\chi^2 = 4.29$; $p = .75$).

All tests of the experimental hypotheses were performed upon the last five judgments (21–25), the means for the individual *Ss* providing the raw

TABLE 4

ANALYSIS OF VARIANCE FOR MEAN OF JUDGMENTS 21–25

Source	*df*	Mean Square	*F*
Serial position	1	.00	
Value of anchor	1	10.01	83.42**
Nature of stimulus	1	4.57	38.08**
Position × anchor	1	2.37	19.75**
Position × stimulus	1	.81	6.75*
Anchor × stimulus	1	.15	1.25
Position × anchor × stimulus	1	.37	3.08
Within groups	120	.12	
Total	127		

* Significant at .05 level.
** Significant at .001 level.

scores. Table 4 presents the analysis of variance based upon these scores. The highly significant interaction between the effects of serial position and the value of the anchor (whether it was high or low) supports the main experimental hypothesis. The subsidiary hypothesis that these effects should be less pronounced with familiar materials does not get full support from the higher-order interaction, which falls somewhat short of the .05 level of significance.

The separate group means are presented in Table 5. According to the main experimental hypothesis, the two high-anchor groups with the anchors in the late position should have smaller mean judgments than the corresponding groups with the early anchors; this is because high anchors tend to push judgments down, and the hypothesis was that this tendency should be greater when the anchor appears in the late position. For the cards, the difference was significant (one-tailed test) at the .05 level ($t = 1.72$); for the men, it was significant at the .01 level ($t = 2.62$). Since low anchors tend to push judgments up, the hypothesis predicts that for the low-anchor groups, serial-position differences will be in the opposite direction. This was very much the case with the cards for which the difference was significant (one-tailed test) at the .0005 level ($t = 4.20$), but it was not the case for the low-anchor men groups between which there was practically no difference.

To test a possible explanation of this lack of difference, two additional low-anchor men groups, with 12 *S*s each, were run. For the new groups, an additional sketch, drawn to appear shorter than any of the other sketches, was added to the set of stimuli; and the "tallest" of the old sketches was removed. This was done because the mean judgment of the sketches was greater than that of the cards throughout the experi-

TABLE 5

MEANS OF JUDGMENTS 21–25*

Value of Anchor	Nature of Stimulus							
	Cards				Men			
	Serial Position				Serial Position			
	Early		Late		Early		Late	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
High	2.86	.36	2.65	.34	3.23	.31	2.91	.37
Low	2.98	.43	3.53	.27	3.69	.28	3.70	.33

* $N = 16$ for each cell entry.

TABLE 6

EXP. III: PRESENTATION OF THE STIMULI*

Groups	Order of Presentation						
	Original Stimuli			Interpolated Anchors			Retest
	1–5	6–10	11–15	16–20	21–25	26–30	31–35
Red—Similar	CBDEA	DCAEB	CBDEA	HGIJF	IHFJG	HGIJF	DCAEB
Red—Dissimilar	CBDEA	DCAEB	CBDEA	hgijf	ihfjg	hgijf	DCAEB
Blue—Similar	cbdea	dcaeb	cbdea	hgijf	ihfjg	hgijf	dcaeb
Blue—Dissimilar	cbdea	dcaeb	cbdea	HGIJF	IHFJG	HGIJF	dcaeb

* Upper-case letters represent red cards; lower-case represent blue cards; letters assigned in order of increasing size.

ment, judgments of the middle sketch tending to be "tall" rather than "medium." This suggested that the sketches portrayed men who were taller than the distribution of real men of *Ss*' experience, and that the resulting asymmetry of judgment masked the serial-position effect. The preanchor judgments of the two new groups were more often "medium," and the difference between their mean judgments of the last five presentations suggested significance in the expected direction (M = 3.47 and 3.80 for early and late groups, respectively; t = 1.86, p = .05, one-tailed test).

EXPERIMENT III

The third experiment was planned to test the hypothesis that a variable known to affect the amount of retention in traditional retroactive-inhibition experiments would also affect the shifts in judgment found by the method of single stimuli. Since retroactive inhibition has been shown to vary with the *similarity* between the original and interpolated activities (3), the present experiment used two degrees of similarity of color between the original stimuli and interpolated anchor stimuli as the experimental variable. The prediction was that the more similar an anchor was to the original stimuli, the greater would be its effect upon rejudgments of these same stimuli.

Method

Materials.—The stimuli were 20 square cards, ranging in width from 3 to 12 in., in steps of 1 in. There was one red card and one blue card for each of the ten different sizes.

Subjects.—There were 40 *Ss* from introductory psychology courses, ten in each of four groups.

Procedure.—The instructions and method of presentation of the stimuli were approximately the same as in Exp. I and II. All stimuli were of one color for the Similar groups. For the Dissimilar groups, the interpolated anchor stimuli—the five largest cards—were a different color. To balance out any effect due to correlation between color and perceived size, two of the groups judged the red cards first, while the other two judged the blue ones first. This design, along with the order of presentation, is shown in Table 6.

Results

As in Exp. I and II, the qualitative size judgments were transformed to

TABLE 7

MEAN DIFFERENCES BETWEEN ORIGINAL AND TEST JUDGMENTS

Test Stimuli	Dissimilar		Similar	
	Mean	*SD*	Mean	*SD*
Red	2.8	2.2	4.0	1.9
Blue	2.2	1.2	3.7	1.3

TABLE 8

ANALYSIS OF VARIANCE OF JUDGMENT SHIFTS IN EXP. III

Source of Variance	*df*	Mean Square	*F*
Similarity	1	18.2	5.7*
Color	1	2.0	
Similarity × Color	1	.2	
Within	36	3.2	
Total	39		

* Significant at .05 level.

numerals. For each *S*, the mean of his five retest judgments (31–35) was subtracted from the mean of his last five judgments before the anchors (11–15), and the statistical tests were performed upon these difference scores. In Table 7 we find support for the experimental hypothesis in the larger mean differences for the Similar groups. This, along with the significant Similarity mean square of Table 8 ($.01 < p < .05$), indicates that the anchors caused the judgments to shift more when the cards were all the same color than when the anchors were of a color different from the original stimuli. Bartlett's test yielded a χ^2 of 4.32 ($p = .25$), permitting the *F* test which is summarized in Table 8. Since neither the Color nor the Interaction mean squares are significant, there is no basis for attributing the main experimental effect to a size-color illusion.

DISCUSSION

The results of these three experiments constitute support for the heuristic value of applying general principles of learning to the analysis of shifts in judgment. While the statistical significance of the results of Exp. I and III is not as high as might be desired, the results as a whole suggest that *S*'s judgment scale and his adaptation level are acquired and modified according to the same principles by which other learning is acquired, retained, or forgotten.

Experiment I demonstrates a positive transfer that might be treated as a form of anchoring; instead of being told by *E* that a certain stimulus defines a judgment category, as in the usual anchoring experiment, *S* is given information from which he can determine for himself the appropriate definition of his judgment categories.

In Exp. II, the results indicate that the order in which stimuli are presented must be taken into account for the prediction of specific judgments. Helson (1, 2) has been able to make successful predictions of judgment, without weighting the stimuli for their serial position, because the stimuli were presented repeatedly in random order, achieving greater stability of judgment (but systematically canceling out any serial position effect).

Although the support for the subsidiary hypothesis of Exp. II, that the serial position effect would be less pronounced with familiar materials, fell short of statistical significance, the possibility of such a limitation should not be rejected. This seems particularly obvious in applications to social psychology, where we would hope that frames of reference concerning enduring social issues would not be at the mercy of the latest anchor.

Experiment III raises more problems than it answers. If variations in the similarity between the stimuli on a nonjudged dimension produce varying amounts of retroactive inhibition, would it be possible to measure this similarity independently of its effects on judgment? This experiment investigated only two levels of similarity, and those two on only one dimension. Under other conditions, modifications of some stimuli in the series (e.g., calling special attention to them) might be expected to increase their anchoring effect. No claim can be made here to an understanding of these phenomena; but it may be, as Pratt (4) argued long ago in conjunction with the time error, that principles of forgetting can be studied better in the relatively simple psychophysical situation than with the more complex materials used in human learning experiments.

Summary

Three experiments investigating the effects of learning variables upon shifts in perceptual judgments were presented.

In Exp. I, the distribution of the stimuli was described to one group of *Ss* before they made their judgments. This group showed less shift in judgment in the course of the subsequent stimulus presentations than did a control group that had not been given the prior description of the stimuli.

In Exp. II, stimuli outside the normal range were introduced either early or late in the series of stimulus presentations. It was found that the anchoring effects of these extreme stimuli upon the final judgments of the series were greater when the stimuli appeared in the later position.

Experiment III was a study of the effects of varying the similarity between the regular and anchor stimuli on a dimension different from the one *Ss* were instructed to judge. Following the conventional retroactive-inhibition paradigm, *Ss* judged as their interpolated activity stimuli that were larger than the original stimuli of the series. For half the *Ss*, the interpolated anchors were the same color as the original stimuli; for the other half, the color was different. Rejudgments of the original stimuli showed a smaller shift in judgment by the latter *Ss*.

References

1. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
2. Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
3. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
4. Pratt, C. C. The law of disuse. *Psychol. Rev.*, 1936, **43**, 83–93.

(Received October 2, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

THE EFFECTS OF A PRE-EXTINCTION PROCEDURE ON THE EXTINCTION OF PLACE AND RESPONSE PERFORMANCE IN A T MAZE [1]

DONALD P. SCHARLOCK

Roosevelt College

Seward and Levy (3) and Deese (2) have recently reported on experimentation dealing with the effects of certain pre-extinction prbcedures on the extinction of (*a*) a simple running response, and (*b*) a U-maze discrimination. Regarding his results Deese points out that once the animals had been trained, the correct response could be extinguished ". . . simply by placing them [the animals] in the goal box without the presence of food" (2, p. 366). In addition, with reference to the results reported by Seward and Levy, he remarks, ". . . a simple non-choice response can be reduced in strength by 'telling' the animals that there is no longer food in the goal box" (2, p. 362). These quotations reflect the underlying hypothesis in both of these experiments, i.e., the hypothesis of nonresponse extinction.

The necessary condition for this type of extinction to occur is that after training a situation be arranged that is adequate to inform *S* of the nonrewarding character of the goal location. If *S* is then subjected to the usual procedure of experimental extinction, the effects of the pre-extinction experience will be such as to facilitate the breakdown of the acquired habit. The effect can presumably be observed by comparing the performance of *S* during extinction with that of control *S*s that do not have the benefit of pre-extinction training.

On the basis of the sign-gestalt theory of learning, it might be anticipated that place and response *S*s trained by the methods outlined by Tolman, Ritchie, and Kalish (6) would exhibit different performances during the extinction session that had been preceded by an opportunity to "perceive" the nonrewarding character of the goal location. If the learning of the place *S* is truly "cognitive," this *S* should benefit more from the pre-extinction training than a response *S*. Accordingly, one might expect the performance of the place *S* to deteriorate more rapidly during extinction.

Since the T maze has, in general, enjoyed greater popularity as an instrument for the study of the acquisition of habits rather than their extinction, a problem arose with regard to the appropriate variable to be used as a measure of extinction.[2]

[1] This article is an excerpt from a dissertation submitted to the faculty of the Graduate School of Arts and Sciences of the University of Buffalo in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is greatly indebted to Dr. B. Richard Bugelski for his assistance, guidance, and encouragement throughout the course of this research.

[2] Seward and Levy (3) used an extinction criterion of two successive failures to leave the starting platform within 3 min. Preliminary observations on the behavior of four *S*s during experimental extinction in the T-maze situation used in this experiment suggested that such a criterion was not suitable. After 60 massed extinction trials *S*s were still taking an average of only 35 sec. to leave the starting position. Furthermore, starting latencies on the later trials of the extinction were not consistently greater than those on the earlier trials. An *S* might leave the starting position on Trial 40 with a latency of only 1 sec. However, in spite of the lack of consistency, there was a general tendency

In view of the sparse treatment of this topic in the literature, it was felt that the problem was essentially an empirical one. Accordingly, observations were made on some five variables that might be presumed to reflect the progress of extinction. These measures will be discussed in detail below.

Method

Apparatus.—In order to bring *S*'s visual environment under *E*'s control, a special enclosure was prepared to contain the T maze. A wooden frame 13 × 13 × 6.5 ft. was erected and covered with unbleached muslin. One corner of the enclosure was arranged to permit the entrance and exit of *E*. From the interior, this corner was indistinguishable from the remaining three. Once inside the enclosure it was quite impossible for *E* to discern any spatial cues other than the maze and the lamps described below.

The maze itself was an elevated T, constructed from 2 × 4-in. white pine, the narrow side being used as the running surface. The starting arm was 4 ft. long and perpendicularly incident to the running arm at its midpoint. It was detachable from the running arm and its position could be varied depending upon whether *S* was required to start from the north or from the south. The over-all length of the running arm was 9 ft. The running surface was elevated 40 in. above the floor. Small metal food cups 1½ in. in diameter and 1 in. deep were affixed to the ends of the running arm, ¼ in. below the running surface. The entire maze was painted flat black and so located in the enclosure that the choice point was directly below the center of the ceiling, with the running arm parallel to one of the sides.

Approximately 6 in. beyond each food cup, and at the same height above the floor, was a 7½-w. frosted, white lamp, mounted on a wooden stand. Hence, if the lamp was lighted and if *S* was going toward the food cup, the lamp was directly in front of *S* and at eye level. The lamp stands were equivalent in structural detail and were painted flat black.

The remaining furniture in the enclosure consisted of a stool for *E* located directly behind the starting arm. During the experimental sessions *S*s were housed in a 20-cell cage located on the floor behind *E* and outside the enclosure. This cage was always located directly behind the starting arm.

During the pre-extinction period a clear plastic shield was used to block the goal arm of the T at a point 11 in. from the food cup. The shield was used in order that *S* might have adequate opportunity to "perceive" the relationship between its position of confinement at the empty food cup and the other features of its environment. A short length of 2 × 2-in. white pine, mounted on a stand outside the enclosure, served as the "neutral platform" to be referred to later.

Subjects.—Sixty albino rats, 33 females and 27 males, ranging in age from 90 to 120 days and having no previous experimental history, were selected from the stock of the animal laboratory maintained by the department of psychology at the University of Buffalo.

Preliminary procedure.—All *S*s were placed on a 24-hr. maintenance schedule 48 hr. prior to the beginning of habituation. Three days of habituation on 2 × 4-in. pine rails set up for that purpose in another room of the laboratory were administered. The *S*s were permitted to run and feed on these rails for approximately 15 min. each day. On Day 4, to examine for position habit, four trials were given to each *S* on the T maze outside of the enclosure with 90° rotation of the maze after each trial. During these trials food was to be found in both food cups. If *S* exhibited a preference for right turns, i.e., three out of its four trials were right turns, it was required to turn left during the experiment proper if it was in the Response Group. Similarly, right turns were required of *S*s exhibiting a left-turn preference. The *S*s in the Place Group either exhibited no particular turn preference, i.e., two turns right and two turns left in the test for position habit, or they were assigned to these groups in such a manner that there were as many right turners as left turners in a given group.

Beginning on Day 5 all *S*s were given four trials a day by the noncorrection method. Seven successive days of training (Days 5–11) were given, with the starting positions ABBA on the odd days and BAAB on the even days. Intramaze cues were varied by rotating the arms of the T around the choice point between trials and by wiping the maze with a damp cloth. In general, training methods followed those described by Tolman, Ritchie, and Kalish (6) with the exception of a longer interval between trials, i.e., at least 10 min. for any given *S*. A recent report by Thompson and Thompson (4) indicated the desirability of increasing the interval between trials since massed practice tends to increase reactive inhibitory factors for the response group, hence placing its performance at a disadvantage.

The *S*s in the Response Group ($N = 30$) were

---

for the starting latencies to increase, although they never tended to approach the level used as a criterion in the Seward and Levy experiment in 60 extinction trials.

TABLE 1

MEANS OF FREQUENCY AND TIME MEASURES DURING EXTINCTION

Group	Frequency Measures		Temporal Measures*		
	Errors	Cup Responses	Starting Latency	Choice-Point Time	Running Time
Place					
Experimental	9.66	18.88	3.60	1.33	13.53
Control	8.66	19.22	3.49	1.26	13.46
Response					
Experimental	9.26	18.53	3.20	1.53	13.06
Control	9.40	18.60	3.80	1.40	13.33

* In estimating the central tendency of the temporal measures for a given *S* throughout the 25 trials, medians were used because of the frequent occurrence of extreme scores.

trained to make a particular turn (right or left depending on their position habits) with the lamps at both ends of the T maze lighted simultaneously. Any given *S* in this group always made the same turn regardless of starting position. The *S*s in the Place Group ($N = 30$) were trained to go to a constant location in the enclosure, half to one end of the maze and half to the other, depending on the position of the single lighted lamp. These *S*s were required to go to the lighted "place" regardless of the turn involved.

Pre-extinction training.—On the last of the seven days of training (Day 11) the two main groups were each divided into experimental and control subgroups. The *S*s were assigned at random to these subgroups. On the next day (Day 12) the plastic block was placed on the maze and pre-extinction training was given to all *S*s.

In the experimental groups all *S*s were confined at the empty food cup four times. They were allowed to remain at the food cup for 1 min. each time, as in Deese's (2) procedure. These trials were spaced 1 min. apart.

In the control groups all *S*s were placed four times on the "neutral platform" outside of the enclosure. They were allowed to remain on this platform 1 min. each time. These trials were spaced 1 min. apart.[3]

Extinction training.—Immediately following the pre-extinction on Day 12, the extinction proper was begun. The plastic block was removed from the maze and each *S* was given 25 trials spaced 4 min. apart. The order of starting positions was ABBA on the first four trials, BAAB on the second four, etc.

Extinction measures.—Observations were made on five variables which, it was presumed, might reflect the progress of extinction. Two frequency measures and three temporal measures were taken.

The frequency measures were: (*a*) *Errors:* The total number of errors made by each *S* during the 25 extinction trials was recorded. An error was made whenever all of *S*'s body crossed onto the incorrect goal arm farther than its own body length. (*b*) *Cup responses:* The total number of cup responses made by each *S* during the 25 extinction trials was observed. A distance 6 in. from each food cup was marked off on the goal arm of the maze. If during a particular trial *S* entered this area and "looked" into the food cup, such behavior was recorded as a cup response.

The temporal measures were: (*a*) *Starting latencies:* A mark was placed on the starting arm of the T, 12 in. from the end. The *S*s were placed on the starting arm in such a manner that they faced away from the choice point. The time required by each *S*, after being placed on the starting arm, to turn around, run toward the choice point, and pass out of this 12-in. area was recorded. (*b*) *Choice-point time:* Marks were symmetrically placed on the maze 2 in. to either side of the choice point on the goal arm and 9 in. back from the choice point on the starting arm. This area was designated as the choice-point area and the time spent by each *S* on every trial in this area was recorded. (*c*) *Running time:* The total time required to traverse the maze from start to goal was recorded for each *S* on every trial.

RESULTS

Table 1 summarizes the observations made on the behavior of all *S*s during the extinction session. The

[3] A recent paper by Bugelski, Coyer, and Rogers (1) has been critical of the design of both the Seward and Levy study and the Deese study in terms of their lack of appropriate control groups.

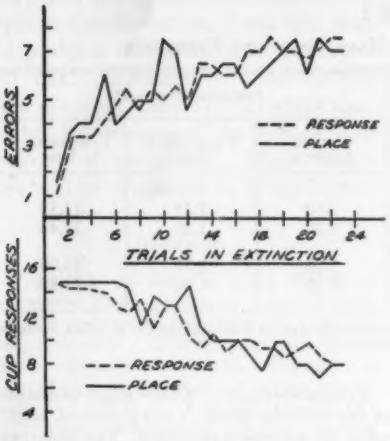


FIG. 1. Distributions of the frequency measures over the 25 trials of extinction

means are for groups of 15 *Ss*. The *SD*'s for each of the measures, estimated on the basis of the within-groups variance, were as follows: (*a*) errors: .89; (*b*) cup responses: 2.71; (*c*) starting latency: .93; (*d*) choice-point time: .65; (*e*) running time: 5.51.

An analysis of variance applied to the raw data on which the tabled means were based showed that no significant differences existed among any of the groups on any of the five measures. However, it was felt that a knowledge of the distribution of each variable throughout the 25 trials of extinction would provide some basis for deciding which of these measures seemed to be most appropriate for the T-maze situation. The distributions of these variables are graphically represented in Fig. 1 and 2. Since the analysis of variance had indicated that no differences existed between the Place Experimental and the Place Control Groups nor between the Response Experimental and the Response Control Groups these data were combined to yield single Response Group and Place Group plots.

The results presented in the above figures may be summarized in the following manner: The number of errors made during extinction of both place and response performance tended to an asymptotic value of 7 by Trial 20. On the basis of the binomial model the mathematical expectation of the number of errors under these conditions is 7.5. In fact, the median number of errors made by the four groups during the first four trials of the original learning was 7. In the case of the cup response measure, practically all *Ss* in both Place and Response Groups investigated the food cup during the first seven trials of extinction. After Trial 7 this behavior began to drop out until, around Trial 20, only about one-half of the *Ss* in each group were engaging in such behavior.

The starting latencies show a general increase over the extinction period from about 1 sec. on the first trial to approximately 8 sec. by Trial 18. The median starting latency for the four groups of *Ss* during the first four trials of their original learning was 19

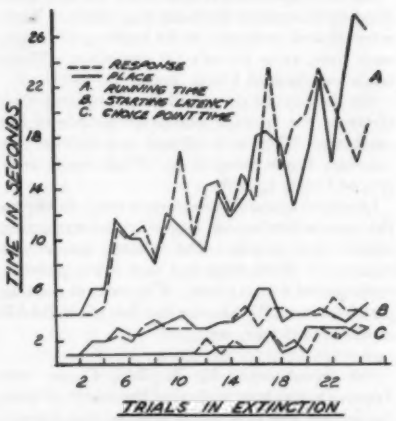


FIG. 2. Distributions of the temporal measures over the 25 trials of extinction

sec. Although starting latencies do tend to increase during extinction, they do not appear to be approaching the level characteristic of the first four trials of the original learning situation. During the extinction trials the time spent by the two groups in the choice-point area does not exhibit much variability, the range being from about 1 sec. on the first ten trials to 3 sec. on the last five trials. The median time spent by the four groups of *Ss* in the choice-point area during the first four trials of the original learning situation was 29 sec. Although the efficiency of the performance in terms of the error measure is rather low by Trial 18, the time spent in the choice-point area has not shown much increase. The total running time for both groups varies from about 4 sec. on the first trial to a level somewhere around 18 or 20 sec. by Trial 20. The median total running time for both groups during the first four trials of the original learning was 57 sec.

Discussion

The results obtained in this experiment fail to support the hypothesis of nonresponse extinction since they afford no basis for assuming that the pre-extinction training, in the case of the two experimental groups, was adequate to convey any information regarding the nonrewarding character of the goal location. These findings are in essential agreement with those reported by Bugelski, Coyer, and Rogers (1), who were unable to verify the Seward and Levy (3) findings and found no evidence for the hypothesis of nonresponse extinction in a situation of their own design. Failure to find differential effects of the pre-extinction training on the performance of place and response animals in this situation may have been a function of the pre-extinction training itself, in the sense that such training was not extensive enough to allow the effect to become evident. However, on the basis of the Deese (2) report, there was no reason to suspect that the amount of pre-extinction training was not adequate. Specifically, no evidence was found in this experiment that the learning of the place animal was in any sense "cognitive." In both the Place and Response Groups, the cup response was observed to drop out early in extinction. In spite of the fact that they did not investigate the food cup, *Ss* continued to leave the starting position with relatively low latencies and execute the appropriate running response with its associated turn. Viewed from the framework of a theory that stresses the importance of "purpose" and "goal orientation," such behavior seems to present certain difficulties of interpretation.

In terms of the thesis of "behavior at the choice point" (5) these data suggest that "decisions" are not made at that particular locus on the T maze. General observations on the performance of *Ss* during extinction, and during training for that matter, tend to support this suggestion. The notion of "behavior at the choice point" may tend to promote a picture of *S* sitting at that location and deliberating on the nature of the choice to be made. Such a picture is hardly consistent with the facts. Characteristically, *Ss* spend very little time at the choice point during any particular run. As a matter of observation, *Ss* exhibit a greater tendency to delay in the pre-choice-point areas of the starting arm, and then proceed without hesitation to make a particular turn. In addition, well-trained *Ss* can be observed to run down the starting arm on the side toward which they will turn.

It will be recalled that no standard experimental procedures have been developed for studying extinction in the T-maze situation and that one of the aims of this study was to compare several possible measures of extinction on an empirical basis. In view of the observations made in this experiment, it would appear that the best measure of extinction of the five investigated was the number of errors. Of all the measures, it seemed to be the only one which, during the course

of the 25 extinction trials, regressed to the level characteristic of the initial stages of learning. Such regression is virtually completed by the eighteenth trial. It would appear that future experiments dealing with extinction on the T maze might be arranged so that the total error scores obtained in 20 trials would serve as a criterion measure.

Summary

An experiment was designed to test the effects of a pre-extinction procedure on the performance of place and response rats during extinction on a simple T maze. The prediction was made that place *Ss* should benefit more from the pre-extinction procedure; i.e., their extinction should proceed more rapidly than that of response *Ss*, if the learning of the place *Ss* is truly "cognitive." A number of different variables was observed in order to determine the measure that would most adequately reflect the breakdown of place and response performance. These included two frequency measures (errors and cup responses) and three temporal measures (starting latency, choice-point latency, and total running time). It was concluded that: (*a*) the pre-extinction training exerted no differential effect on the performance during extinction of place and response animals; (*b*) the thesis of "behavior at the choice point" is at least misleading and at best an oversimplification of T-maze performance; and (*c*) errors seemed to reflect more adequately than any other measures the breakdown of performance during extinction.

References

1. Bugelski, B. R., Coyer, R. A., & Rogers, W. A. A criticism of pre-acquisition and pre-extinction of expectancies. *J. exp. Psychol.*, 1952, **44**, 27–30.
2. Deese, J. The extinction of a discrimination without performance of the choice response. *J. comp. physiol. Psychol.*, 1951, **44**, 362–366.
3. Seward, J. P., & Levy, N. Sign learning as a factor in extinction. *J. exp. Psychol.*, 1949, **39**, 660–668.
4. Thompson, M. E., & Thompson, J. P. Reactive inhibition as a factor in maze learning: II. The role of reactive inhibition in studies of place learning vs. response learning. *J. exp. Psychol.*, 1949, **39**, 883–891.
5. Tolman, E. C. The determiners of behavior at a choice point. *Psychol. Rev.*, 1938, **45**, 1–41.
6. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning. V. Response learning vs. place learning by the non-correction method. *J. exp. Psychol.*, 1947, **37**, 285–292.

(Received October 22, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

MOTIVATIONAL PROPERTIES OF FRUSTRATION: II. FRUSTRATION DRIVE STIMULUS AND FRUSTRATION REDUCTION IN SELECTIVE LEARNING[1]

ABRAM AMSEL AND JOSEPH S. WARD[2]

Tulane University

An experiment by Amsel and Roussel (1) offered evidence of the contribution of frustration to generalized drive strength. Frustration was defined as "a state which results from the non-reinforcement of an instrumental response which previously was consistently reinforced." They showed that adding frustration to the motivational complex increases the strength of an immediately following, hunger-motivated response, presumably through increase in generalized drive strength. Other investigators (8, 10, 11) have offered the same interpretation for similar findings.

If frustration has motivational properties like hunger, pain, or fear, it should be possible to show that it has directive (drive stimulus) properties. In their recent theoretical treatment of frustration, Brown and Farber write that " . . . such [frustration] stimuli could provide the means whereby an organism might learn to discriminate between (react differentially to) frustration and other states such as fear or hunger" (2, pp. 489–490).

To demonstrate this drive stimulus property of frustration, one would have to set up experimental conditions wherein frustration serves as a cue in the learning of a new response or discrimination. A modification of the Hull-Leeper technique (6, 7) provides one method of accomplishing this. It involves training Ss to make one response following frustration and an alternative response following nonfrustration (reinforcement) in the same situation. If such differential response selection can be learned, there would be some evidence for the operation of a frustration drive stimulus.

There is, however, one major difficulty with the above plan: the very operations which define the frustration and nonfrustration conditions may also be thought of as giving rise to differential stimulation of a more peripheral sort. For example, if hunger and food are the motivational and reward conditions, differential stimulation might arise from food-in-mouth as against no-food-in-mouth, apart from any frustration induced.

The experiments which follow are an attempt to identify the motivational (drive and drive stimulus) properties of frustration by showing that: (*a*) selective learning can develop rapidly on the basis of reinforcement and nonreinforcement as differential cues; (*b*) factors that presumably would reduce the effectiveness of peripheral (mouth) cues do not retard the learning; (*c*) responses followed by frustration reduction tend to be fixated compared with those that are not followed by frustration reduction.

Before proceeding to a description of the experiments, we must come back to our earlier definition of frustration. The procedure to be followed here is generally comparable to that employed by Amsel and Roussel (1): A

[1] This study was supported by a grant from the Tulane University Council on Research.

[2] Now at Human Resources Research Office, Washington, D. C.

series of trials is run in which *S* is consistently rewarded for a particular response and is then immediately allowed to make a second similar response for which it is again rewarded. Each trial, then, consists of two consecutive rewarded responses. There follows another series of trials in which the first of the two responses is rewarded on half of the trials but not on the other half, and, in either case, a second response is made. In the Amsel and Roussel situation, the vigor of the second response was measured as an indicator of increased drive strength attributable to frustration following the nonreinforcements as compared to the reinforcements. In the following experiments, after a series of double reinforcements, *S* is required to make one response (e.g., left turning) following reinforcement of the first response, and another response (e.g., right turning) following nonreinforcement of the first response. It is clear that after the first nonreinforced trial, subsequent nonreinforcements cannot be said to follow "consistent" reinforcement, if by "consistent" is meant reinforcement on every previous trial. However, the Amsel and Roussel experiment demonstrated that following a long series of reinforcements, the "frustration effect" persists on the nonreinforced trials of 50% reinforcement (test) conditions. Indeed the effect has been shown to develop without prior 100% reinforcement under 50% reinforcement conditions (**10**). In our present experiments, therefore, all nonreinforced response during the test (discrimination) period will be considered as frustrated response, because the conditions for the "frustration effect," demonstrated earlier, are present.

Apparatus

The apparatus was in two parts, A and B. Apparatus A consisted of: a gray starting box, 12 in. long and 4 in. wide; a straight, gray runway, 42 in. long and 4 in. wide; a black goal box (G_1), 12 in. long and 4 in. wide; a second gray runway, 15 in. long and 4 in. wide; and a second black goal box (G_2), 12 in. long and 4 in. wide—all connected in series. The starting box and first runway were separated by a guillotine door manipulated by *E*. The runway and G_1 were separated by a swinging door, which opened into the goal box but not in the reverse direction. (This section became the stem of a single-choice T maze during discrimination trials.) G_1 was separated from the second runway by a guillotine door manipulated by *E*. The second runway and G_2 were separated by a swinging door, which opened into the goal box but not in the reverse direction.

Apparatus B was a single-unit T maze. It consisted of the stem portion (starting box, runway, and G_1) of Part A, and a cross portion to complete the T. There were two cross portions employed in the experiments, completing two different T mazes. The cross portion of one T (B–W) was divided into black and white arms. Leading to the right of the choice point, following G_1, was a 15-in. white alley leading into a white goal box, 12 in. long and 4 in. wide. To the left of the choice point were an alley and goal box, both painted black. Alleys and goal box were separated by swinging doors, which opened into the goal box but not in the reverse direction. The second T (B–B) was identical in size and structure to the first. The only difference was that both end boxes were black and were separated by a gray alley.

A 1-in. square of galvanized iron, painted the color of the goal box in which it was placed, was used to hold the drops of water or pellet of food on the reward trials. This was removed on nonreward trials.

The apparatus was on a table 28 in. high. The *E* stood behind a wire mesh screen which was painted white on the apparatus side. Illumination was provided by two 40-w., white fluorescent strips suspended 28 in. above the apparatus in the shape of a T.

General Procedure

For about two weeks prior to the preliminary training, *S*s were handled daily and adjusted to the appropriate drive regimen. The particular schedules of deprivation employed will be described separately with each experiment.

The entire experimental sequence for each experiment was in five stages:

1. In addition to being handled and becoming adjusted to the deprivation schedule, *S*s were trained to run through swinging doors. Each *S* was given six trials in two days, and by the end of this period all *S*s readily manipulated the

doors. This period lasted two days in all experiments, and on each day *S* was rewarded three times for pushing through the swinging door.

2. The *S*s were placed in the starting box of Apparatus A. The *E* raised the guillotine door and allowed *S*s to explore the runway, to push through the swinging door, and to receive the reward in G_1. Ten trials were spread over three days of this period, two on the first day and four on each of the next two days.

3. In all experiments, the procedure was the same during this third stage of training. Each *S* ran five trials per day in Apparatus A. The *S* was placed in the starting box, and *E* raised the guillotine door, allowing *S* to traverse the first runway, raise the swinging door, and enter G_1. After *S* finished eating (drinking), *E* raised the second guillotine door, allowing *S* to traverse the second runway and enter G_2. The *S* was again rewarded in G_2. Thirty-five trials were run in this manner.

4. The fourth stage of the procedure was the period during which the actual discrimination training was carried out. Since the separate experiments differ in certain details of procedure during this period, this part of the procedure will be described separately for each experiment.

Experiment I

The first experiment was performed to determine whether or not hungry rats could learn, in a T maze, to make one response at the choice point (turn left) following food reinforcement, and the spatially opposite response (turn right) following nonreinforcement (frustration). More specifically, if *S* is given food in a goal box just before the choice point (G_1), a left-turning response leads to further reinforcement in the goal box (G_R); if, however, *S* finds no food in the choice-point goal box, a right-turning response must be made when *S* is released, in order for food to be available in the goal box (G_F).

Experimental conditions.—The *S*s for Exp. I were five male hooded rats about 120 days old. All *S*s were deprived of food for 22 hr. at the start of each day's trials.

Apparatus A and Apparatus B (B–W) were used. In the T maze, a left turn led into a black alley and goal box, and a right turn into a white alley and goal box.

Prior to discrimination training, each *S* was reinforced 80 times in the two black goal boxes of Apparatus A as described earlier. On the first day of discrimination training, *S*s were first given two trials in Apparatus A, followed by four trials in Apparatus B. On this day they were forced to run twice to the left following reinforcement and twice to the right following nonreinforcement in G_1. On the 33 subsequent discrimination days, four trials were run in Apparatus B, preceded by a single trial in A. On all trials in B, after the first day, *S* was either reinforced or nonreinforced in G_1 according to a prearranged order. Whether or not a particular choice-point response led to further reinforcement depended upon the response of *S*. Noncorrection was employed throughout.

When *S* found food in G_1, it was released after consumption of the food. The time taken to eat the food was usually 10 to 15 sec. On trials when *S* found no food in G_1, the interval in G_1 was varied randomly between 10 and 15 sec. to insure that a temporal differentiation could not be established. Successive trials for each *S* were separated by about 7 to 8 min. The order of reinforcements and frustrations in G_1 was RFRFFRFRRFFRFRRF.

Results and discussion.—The results are shown graphically in Fig. 1. One curve shows the percentage of correct responses on successive blocks of trials. Each point represents percentage correct out of 60 trials (12 trials each for five *S*s). The other curve (dashed line) shows, at each point, the percentage of the trials, correct or incorrect, out of 60, run to the white goal

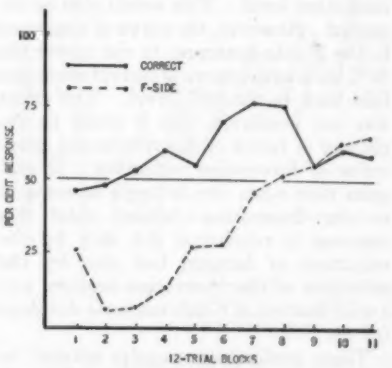


Fig. 1. Learning curve and curve of responses to the frustration (white) side for five *S*s of Exp. I

box, which was the side correct following frustration (F side).

The *S*s as a group reached the 75%-correct response level by about Trial 85, after which the level of correct response declined to about 60%, at which time training was discontinued. Three of the five *S*s reached a criterion of ten successive correct responses—two of the three reached this criterion twice. All *S*s, including the three who reached the criterion of learning, were performing at about the chance level when training was terminated. These data suggest: (*a*) that the discrimination can be learned, and (*b*) that, once learned, it is very unstable and is quickly lost.

Another very interesting feature of the results was noted and is represented in the dashed-line curve—that is, all of the *S*s very quickly developed a preference for the response into the black alley and goal box (R side). This can be seen from the fact that the curve for responses to the F side drops very quickly to almost zero. This might have been predicted on the basis of the secondary reinforcing properties of black, which would presumably arouse r_g-s_g at the choice point leading to the left-turning response. The curve then rises to about 50% as the curve for correct responses reaches its maximum level. This would also be expected. However, the curve of responses to the F side continues to rise above the 50% level as the curve of correct responses falls back to the 60% level. This effect was not predicted, but it could be explained in terms of the reinforcing properties of frustration reduction. It suggests that when the *S*s begin to respond to the frustration (white) side, this response is reinforced not only by the reduction of hunger, but also by the reduction of the frustration tension, and a mild fixation of F-side response develops (cf. 5).

These preliminary results seemed to indicate: (*a*) that the discrimination was partially learned on the basis of reinforcement and nonreinforcement as the discriminanda, but that it was unstable, and (*b*) that there was, in the latter stages of training, a tendency for *S* to make a greater than chance number of responses to the side correct following frustration (F side). As a tentative explanation, this mild fixation might be attributed to double reinforcement on frustration trials: hunger reduction and frustration reduction.

If it could be established with greater confidence that the factors suggested above are operating, it would increase the plausibility of ascribing selective learning to the presence of a frustration drive stimulus on nonreinforced trials, rather than to the presence of more peripheral sources of peripheral stimulation, i.e., mouth cues.

Experiment II

In the second experiment, an attempt was made to minimize the factor of mouth cues in the selective learning situation. The following changes in the general procedure were instituted: (*a*) the reward was changed from a pellet of food to two drops of water at room temperature (about 78° F.); (*b*) on trials in which *S* was rewarded in G_1, 2 to 5 sec. were allowed to elapse between drinking of water and release of *S* to the choice point for the next response. It was felt that this would greatly reduce any difference in the stimulus trace pattern that might arise from water-in-mouth as against no-water-in-mouth.[3]

[3] Smith (12) trained rats to turn left after passing through a black alley in the stem of a T and to turn right after passing through a white alley. The choice-point response was delayed 0, .6, .75, 2.0, or 5 sec. in five groups. Using hooded rats of the same strain employed in the present study, he found that with a 2-sec. delay, the median number of trials to a criterion of 18 correct out of 20 was 580; with a 5-sec. delay, the median was 1,400 trials. With zero delay, his hooded *S*s took 70 trials to the criterion. Smith attributed this sharp increase in trials to learn with increased delay to a sharp decrease in the intensity of the visual stimulus trace.

This modified procedure was carried out under two conditions. Condition B–W (black and white arms on the T) was run as a check on the strong, early fixation of the R-side response and the final predominance of F-side responding shown in Exp. I. Condition B–B (G_R and G_F, both black and separated by gray alley) served as a control for the black-white factor in Exp. I. In this condition, there should be no differential secondary reinforcement effect in the two end boxes and no differential r_g–s_g effect in the alleys leading to the end boxes. It was our purpose to determine, by a comparison of these two conditions: (*a*) whether or not the early tendency to fixate the R-side response would be eliminated, and (*b*) whether or not the final preference for the F-side response would still be manifested in Group B–B.

If the early fixation of the R-side response, shown in Exp. I, were due to secondary reinforcement effects, this effect should not appear in Condition B–B, and the early discriminative performance of Group B–B should be better than that of Group B–W; and if the final preference for the F-side response is due to frustration reduction factors, and is not related to avoidance of black and/or to the stimulus properties of white, it should occur in both groups.

Finally, if the level of selective performance in this experiment is as high as, or is higher than, that in Exp. I, the discriminative performance cannot easily be accounted for solely in terms of mouth cue factors.

The drive condition of the *S*s was determined by 21 hr. of water deprivation and by the ingestion, just before each day's trials, of a maximum of 4 gm. of dry Purina checkers. After each day's runs, each *S* was given 8 gm. of Purina with water for 1 hr.

Following 35 trials of double reinforcement in Apparatus A, as described earlier, the discrimination training began. On every day of this training, 14 trials were run: the first 2 trials were in Apparatus A, and the next 5 were in the T maze; then 2 more trials were run in A and 5 more in B. So, on each day, there were 10 discrimination trials and 4 additional double reinforcements in the black goal boxes of A to maintain the frustrational effect of nonreinforcement.

On all reward trials in the T maze, two drops of water were placed on the metal square in G_1 and in G_R; on all frustration trials, no water was in G_1, and two drops of water were in G_F.

After 200 trials (20 days) of training, a two-day period was introduced during which *S*s ran, on each day, all 14 trials in A. During these days, no trials were run in the T maze. This procedure was designed to increase "reward expectancy" in G_1 if it were not still high. There followed five more days of training in the T maze as before. In all, there were 250 trials of discrimination training.

On all reward trials, *S* remained in G_1 and drank two drops of water. Following cessation of drinking, *S* was delayed from 2 to 5 sec., at random, before *E* raised the guillotine door allowing it to go to the choice point. If *S* went to the left (G_R) after having received water in G_1, a further reward of two drops of water was given; if *S* went to the right (G_F), there was no further reward.

On nonreward trials, *S* found no water in G_1. However, *S* was confined in G_1 for the same length of time as on the previous rewarded trial. The *E* then raised the guillotine door to the choice point. If *S* went to G_F, a reward of two drops of water was given, and *S* was removed on completion of drinking. If *S* went to G_R on any nonreward trial, no reward was given.

The order of presentation of reward and nonreward in G_1 was RFRFFRFRRFFRFRRFR-FRF. Ten *S*s were run at a time, and in rotation, so the time between trials was about 10 min. for each *S*.

The *S*s were 20 naive hooded rats—10 in each group. There were 7 females and 3 males in each group, and the age range at the start of the experiment was about 120–150 days.

Results.—Figure 2 shows the total number of correct responses for Groups B–B and B–W on each day of the experiment. Both groups show selective learning on the basis of reward or frustration in G_1. The groups were

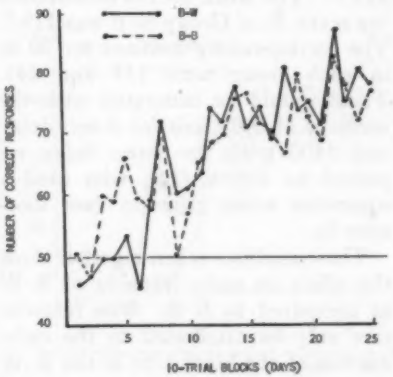


FIG. 2. Learning curves for two groups, with differential alley and end-box stimulation in one (B–W), but not in the other (B–B)

maintaining a level of about 80% correct responding at the end of the 25-day period.

In order to test whether or not there was an early retardation in the learning of Group B–W relative to B–B, the number of trials taken by each *S* to reach two different criteria of learning was calculated. One criterion—an "easy" one—was 8 out of 10 trials correct on a single day; the second criterion—a more rigorous one—was 19 out of 20 trials correct, with the last 10 correct. Each *S* in Group B–B reached the first criterion before its corresponding *S* in Group B–W. A *t* test of the mean number of trials to reach the criterion showed a significant difference between the groups beyond the .01 level of confidence ($t = 3.57$, $df = 9$).

When the more rigorous criterion of learning was used, only six *S*s of Group B–B reached it before their partners in Group B–W. Two *S*s in Group B–W and one in Group B–B failed to reach this criterion. If we exclude the pairs of *S*s in which one failed to reach the criterion, the mean number of trials for the seven *S*s of Group B–W to reach the criterion was 118.7. The mean for the corresponding seven *S*s of Group B–B was 119.7. The corresponding medians for 10 *S*s in each group were 135 and 145. These should be compared with the medians of 600 trials for 2-sec. delay and 1400 trials for 5-sec. delay reported by Smith (12), who used a somewhat easier criterion (see footnote 3).

These analyses taken together show the effect on early learning of B–W as compared to B–B. The retardation may be attributed to the early fixation of the black side in the B–W group. A comparison of the total correct responses for pairs of *S*s from Days 1–7 also shows significance beyond the .01 level of confidence ($t = 3.62$, $df = 9$). The same analysis for Days 8 to 25 shows no statistically significant difference.

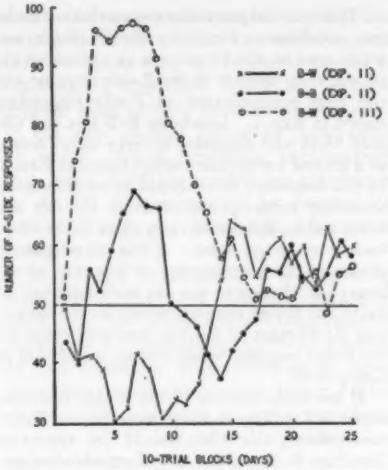


FIG. 3. Comparison of number of responses to the F side for the two groups of Exp. II, and one group of Exp. III

The curves presented in Fig. 3 contribute additional evidence for considering that the learning is based upon the presence of a frustration drive stimulus as one of the discriminanda. They represent the total number of responses to the side correct following frustration (F side) for each day of the experiment. It should be remembered that responses to the F side may be either correct or incorrect. The curve for Group B–W shows the rapid initial fixation of the black side (low level of response to the F side) until Day 12, followed by a rise to about 60% response to the F side, which was maintained for the last 11 days. The *S*s in Group B–B show a very fast rise in level of response to the F side from the first to the 90th trial. From this point, as the curve for correct response rises, the F-side response curve falls to a level below 50%, which is followed in turn by a rise of the F-side curve to about the

58% level, where it remains until the end of the experiment. The difference between the two curves for the first 13 days was significant beyond the .05 level of confidence ($t = 2.34$); the difference between the curves for the last 12 days was not statistically significant.

To test the significance of the apparent mild fixation of the F-side response at the end of training in both groups, the following was done: a score was determined for each *S*, which was its number of responses to the F side on the last five days (50 trials). For Group B–W, the mean of these scores was 29.3, for Group B–B the mean was 28.0. The performance of both groups, at this stage, was about 80% correct. If the incorrect responses were equally divided between the F and R sides, the mean response frequency to the F side should be 25. Taking 25 as the hypothetical mean, t for Group B–W was 1.51 ($.20 > p > .10$), and t for Group B–B was 1.96 ($.10 > p > .05$). Combining the groups, t was 2.28 ($.05 > p > .01$, $df = 19$).

Out of the 20 scores of the combined groups, 16 were greater than 25, 3 were less than 25, and there was one score of 25. A chi-square analysis was performed taking 19 as total frequency, and hence 9.5 as the expected frequency. Chi square was 8.89 ($p < .01$). It would therefore appear that the combined groups were responding beyond the chance level to the F side on the last five days of discrimination training.

Experiment III

Experiment III was a control for the factor of B–W in these experiments. In Exp. I and II, the side appropriate to frustration was white when the B–W condition was employed. Experiment III is a repetition of Exp. II (B–W), except that the sides are reversed: the F side is black, the R side white. If our previous interpretations are tenable, we should find (*a*) that the initial fixation of the black (left) side is even stronger than before because of a combination of secondary reinforcement and frustration reduction factors, and (*b*) that there is still, after learning, a better than chance level of responding to the F side, which is now black.

Experimental conditions.—The procedure in Exp. III was identical with that of Exp. II, except for the reversal of B and W. The *S*s were 10 naive hooded rats, 7 male and 3 female, about 150 days old.

Results.—The results are shown in Fig. 3 and 4. In Fig. 4, the learning curve for the present group is shown in comparison with that of the B–W group from Exp. II. These curves are highly similiar in general form and in the level to which they rise. Figure 3 compares Group B–W (F_B) of this experiment to the two groups of Exp. II with respect to their level of responding to the F side. Group B–B of Exp. II had, at first, fixated the right side, which was the F side. Group B–W (F_W) of Exp. II had fixated the left side, which was not the

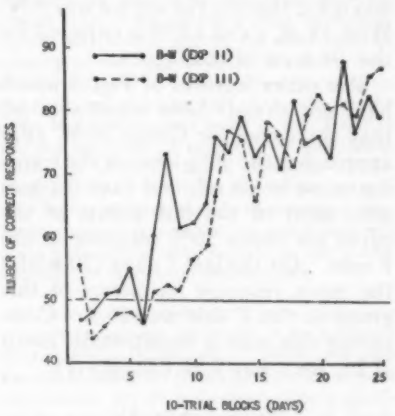


FIG. 4. Comparison of learning of B–W groups where F side is white (Exp. II) and where F side is black (Exp. III)

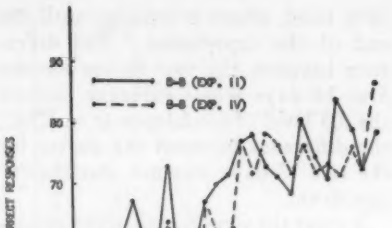
F side but was the black side. Group B–W (F_B) of the present experiment reached a level of response to the left (black) F side of almost 100%. In this group, we have combined both the factors leading to early fixation of one or the other side that were separated in the two groups of Exp. II. In group B–W (F_W) we have, apparently, the secondary reinforcing properties of black outweighing the effect of additional reinforcement from frustration reduction in the early stages of training; in Group B–B, where no differential secondary reinforcement is operating, the F side is responded to, predominantly, during the early trials; in Group B–W (F_B), both factors—differential secondary reinforcement and frustration reduction—operate to the same side, and this side is selected almost 100% at one point during the early trials.

In order to test the significance of these differences, a score was determined for each *S*, which was his number of responses to the F side in the first 100 trials. Using these scores, Group B–W (F_B) was compared with Group B–W (F_W) and with Group B–B. The *t* for the first comparison was 6.85; that for the second was 5.31. With 18 *df*, a *t* of 2.878 is required for the .01 level of confidence.

The other features of Fig. 3 which have not already been noted are: (*a*) that the curve for Group B–W (F_B) approaches the 50% level as the learning curve levels off, and that (*b*) here also, most of the last points of the curve are above 50% response to the F side. On the last 5 days (50 trials) the mean response frequency of this group to the F side was 26.8. Comparing this with a hypothetical mean of 25, the *t* was 1.51 ($.2 > p > .1$).

Experiment IV

Experiment IV was a further control for Exp. II. It was run to provide an additional check on the degree to which the selective learning was based upon mouth stimulation. In this experiment, *S*s were run under "predrinking" conditions. The procedure of Exp. II was exactly repeated, except for one change: on every trial, each *S* was given two drops of water in the starting box of Apparatus A or B so that when *S* reached the first goal box, it had just finished drinking: the time to run the short runway between the starting chamber and the first goal box varied between .5 and 2 sec.



Exp. 5. Comparison of learning of B–B groups with (Exp. IV) and without (Exp. II) predrinking

It was reasoned that if, in fact, the learning depends upon water-in-mouth vs. no-water-in-mouth as the discriminanda, this group should learn significantly slower than the corresponding group of Exp. II, since on every trial, whether *S* is reinforced in the choice-point goal box or not, there should be traces of water present from drinking in the starting box.

On the other hand there is evidence (**3, 4, 9**) that prefeeding—and presumably also prewatering—increases "reward expectancy." If this is so, we should expect, at least, no impairment in the learning, since frustration

presumably depends upon the anticipation of reward.

This experiment was run under the B–B condition, so the results are most comparable to the B–B group of Exp. II. The *S*s were 10 naive hooded rats, 7 male and 3 female, about 150 days old.

Results.—Figures 5 and 6 give the comparisons between the results of this experiment and the B–B group of Exp. II. Examination of Fig. 5 reveals practically complete overlapping of the learning curves of the two B–B groups. Clearly, the giving of water in the starting box did not impede the learning. Figure 6 gives the picture of responses to the F side. The present B–B (predrinking) group shows an even stronger tendency to this response than does our original B–B group. Out of 25 points on the curve, our Exp. IV group was above 50% response to the F side on 24 trials, although it did show the initial rise and subsequent dip of our Exp. II curve. Examination of the raw data indicates that this is not simply the perpetuation of an original position preference; out of the first two trials for the 10 *S*s (20 trials in all), 11 were to the F side and 9 to the R side; out of the first 30 trials, the numbers are 15 to one side, and 15 to the other. The last 5-day (50-trials) analysis shows a mean response to the F side of 28.9 which was not, however, significantly different from chance ($t = 1.35$, p about .2).[4]

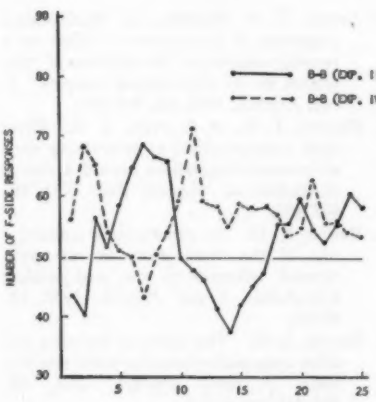


FIG. 6. Comparison of responses to the F side for the B–B groups with (Exp. IV) and without (Exp. II) predrinking

EXPERIMENT V[5]

One last control experiment will be reported. This is again a control for the B–W conditions. The question that might be asked is this: When *S*s are run in a choice situation involving black and white alleys branching from a black stem goal box, should one expect that eventually the tendency to run to the black side will be weakened and that to the white side strengthened because, perhaps, of something in the nature of the stimulation? Could this account for the switch-over in the B–W groups of Exp. I and II?

Five naive, female, hooded rats, about 120 days old, were run under the B–W condition, with the modification that the problem was made insoluble: on one-half of the trials following reward in G_1, *S* was rewarded again for running to the left (black); and on the other half, for running to the right (white). Similarly, frustration in G_1 led to reward half of the time on the right, and the other half of the time on the left. If the switch-over found in Exp. I and II was due simply to the nature of the stimulation employed, we should find it here also. If not, we might expect that there will develop a tendency to run to black because of secondary reinforcing properties, and that no switch-over will occur in the group curve because no side is consistently the frustration-reducing side.

[4] An over-all analysis, not reported under any separate experiment, was made of the combined data of Groups II, III, and IV, which have comparable procedures. In the last 50 trials of these experiments, the 40 *S*s involved made a significantly better than chance number of F-side responses ($t = 2.88$, $\chi^2 = 11.9$, $p < .01$ in both cases).

[5] The authors are indebted to Miss Joan Eversmeyer, who assisted in collecting the data of this experiment.

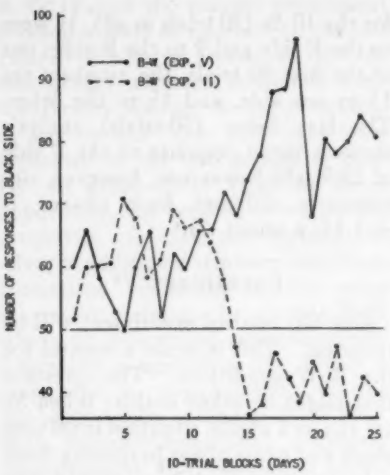


FIG. 7. Comparison of responses to black in two B–W groups where learning is possible in one (Exp. II), but not in the other (Exp. V)

Results.—Figure 7 shows a comparison of the present B–W control with the B–W curve of Exp. II insofar as running to black is concerned. The difference is obvious and unmistakable. Both curves start up together because, in effect, the two conditions are equivalent in the early stages of learning. However, at about the point where the Exp. II group starts to learn, its level of response to black drops, following which the white side (F side) is responded to predominantly. The control curve continues to rise, going up to 96% response to black on Day 19. The difference between the curves on the last 15 days of training is significant beyond the .01 level of confidence ($t = 3.90$, $df = 13$). On the first 10 days the difference is not statistically significant ($t = .32$, $df = 13$).

SUMMARY

This investigation concerned certain implications of ascribing motivational properties to frustration (frustrative nonreward). Using 50 hooded rats as *S*s, and a modified selective learning technique, five experiments were performed. The findings of these experiments may be summarized as follows:

1. Under certain circumstances, reward and nonreward can serve as discriminanda for selective learning. This learning seems to develop rapidly relative to comparable learning involving differential exteroceptive stimulation.

2. Factors that might be presumed to reduce the peripheral cue aspects of the reward and nonreward discriminanda (food-in-mouth vs. no-food-in-mouth), such as employing water as the rewarding agent and delaying the choice-point response to reduce the effectiveness of the stimulus trace of water-in-mouth, do not retard the learning.

3. The response which is correct following nonreward (the F-side response) tends to be made more frequently than that which is correct following reward. This shows up both early and late in the training sequence.

4. The degree to which the F side (or R side) is fixated early in the training sequence is related to the similarity of the F side (or R side) to the choice-point goal box in which *S* has been rewarded. The fixation of the F side late in training seems unrelated to this factor.

According to our interpretation, these findings constitute further support of the hypothesis that frustration, as here defined, has motivational properties. More specifically, they support the hypotheses (*a*) that frustration provides drive stimulation which gives it directive properties, and (*b*) that frustration reduction is reinforcing.

REFERENCES

1. AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.
2. BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.* 1951, **48**, 465–495.
3. BRUCE, R. H. An experimental investigation of the thirst drive in rats with especial reference to the goal-gradient hypothesis. *J. gen. Psychol.*, 1937, **17**, 49–60.
4. BRUCE, R. H. The effect of lessening the drive upon performance by white rats in a maze. *J. comp. Psychol.*, 1938, **25**, 225–248.
5. FARBER, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.

6. Hull, C. L. Differential habituation to internal stimuli in the albino rat. *J. comp. Psychol.*, 1933, **16**, 255–273.

7. Leeper, R. The role of motivation in learning: a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.

8. Marzocco, F. N. Frustration effect as a function of drive level, habit strength and distribution of trials during extinction. Unpublished doctor's dissertation, State Univ. of Iowa, 1951.

9. Morgan, C. T., & Fields, P. E. The effect of variable preliminary feeding upon the rat's speed of locomotion. *J. comp. Psychol.*, 1938, **26**, 331–348.

10. Roussel, J. Frustration effect as a function of repeated nonreinforcements and as a function of the consistency of reinforcement prior to the introduction of nonreinforcement. Unpublished master's thesis, Tulane Univ., 1952.

11. Sheffield, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.

12. Smith, M. P. The stimulus trace gradient in visual discrimination learning. *J. comp. Psychol.*, 1951, **44**, 154–161.

(Received October 26, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

STUDIES IN INCIDENTAL LEARNING: I. THE EFFECTS OF CROWDING AND ISOLATION

LEO POSTMAN AND LAURA W. PHILLIPS

University of California

This paper reports the first of a series of experiments concerned with the conditions and characteristics of incidental learning. The fact that *Ss* will learn even when they are not explicitly instructed to do so has been repeatedly demonstrated (**6**, pp. 210 ff.), but only little is known about the variables on which amount of incidental learning depends and about the properties of associations that have been formed under incidental conditions. We may profitably begin with the question whether or not the variables that are known to influence intentional learning govern incidental learning as well.

Our first experiments compare the intentional and incidental learning of "crowded" and "isolated" materials. Incidental learning is, almost by definition, selective. Degree of perceptual isolation has been claimed as an important determinant of selectivity in learning and retention (**3, 5, 10**). It has, moreover, been asserted that it is under conditions of incidental rather than intentional learning that perceptual organization becomes a decisive determinant of association. Thus, Köhler suggests that "we may have spontaneous association when organization is spontaneous," and that "intentional association is needed where the material itself is slightly organized" (**4**, p. 286). It follows that items which stand out as figures against a homogeneous background should be favored in incidental learning and, indeed, more so than in intentional learning.

Evidence for the hypothesis that perceptually isolated items are learned and retained better than crowded items was presented by Von Restorff (**10**). When an isolated item, e.g., a number, is embedded in a series of like items, e.g., syllables, the isolated item is recalled much better than the homogeneous items. These results were interpreted as supporting the gestalt theory of memory. The "traces" of homogeneous items tend to assimilate and lose their identity, whereas the traces of the "figural" isolated items remain distinctive.

As Saul and Osgood (**8**) have pointed out, the procedure used by Von Restorff leaves room for serious differences in interpretation. Two variables—the absolute numbers of the two types of items and degree of perceptual isolation—are confounded. Both an interpretation in terms of intraserial interferences and the hypothesis of perceptual isolation would predict superior learning of the isolated item. The two interpretations would, however, differ in their predictions concerning retention. If we assume a process of progressive development in the memory trace, the gestalt hypothesis would have to predict an increasing difference in retention between isolated and homogeneous items over time. Interference theory would predict, on the other hand, that if original learning for isolated and homogeneous items is held constant, the temporal course of retention should be the same for both types of items (**8**, p. 378). This assumes that both crowded and isolated

items have been subject to the same amount of interpolated activity.

To control the absolute frequencies of crowded and isolated items, Siegel (**9**) and Saul and Osgood (**8**) used lists in which two types of stimuli—syllables and numbers—occur equally often but appear with varying degrees of isolation within the series. The series used in their studies as well as in our experiments appear in Table 1. With these materials, Siegel found superior retention of the isolated items in a test of immediate recall. The results of Saul and Osgood show less effect of isolation in immediate memory, nor did these investigators find differential recall for isolated and homogeneous items. They conclude that "isolation and massing are effective only in terms of the number of similar items and resulting possibilities for intraserial interference" (**8**, p. 378).

The question still remains open, particularly in view of Köhler's argument cited above, whether perceptual isolation produces greater effects in incidental than in intentional learning. The following experiments were designed to answer this question.

Experiment I

Method

Conditions of learning.—To expose *S*s to the stimulus lists without instructing them to learn, we adapted a procedure used by Jenkins (**2**). Two *S*s are used at a time, one of whom acts as experimenter (*E*) while the other serves as subject (*S*). Both individuals receive the same exposure to the stimulus materials, but only one of them, *S*, is instructed to learn. The *E* is an incidental learner.

When students appeared for the experimental session, they were first informed that one of them was to act as *S* and the other as *E*. Two reasons for this procedure were given: (*a*) the investigator was in need of assistance because of the variety of operations involved in the experiment; (*b*) experience as *E* would provide useful training for the student.

The stimulus list was exposed on a memory drum. The *E*'s task was to read the syllables and numbers aloud to *S* as they appeared in the window of the drum. The *S* repeated each item as soon as it had been read by the *E*. The *E* and *S* were seated side by side in front of the drum so that both had exactly the same experience with the stimulus items; i.e., they saw the items for the same amount of time, and each pronounced the items once and heard them pronounced once.

The experiment was presented as a study concerned with the effects of visual and auditory presentation on the learning of nonsense materials. The intentional learner (*S*) was instructed to learn as many items as possible, though not necessarily in the order in which they were shown. The incidental learner (*E*) was instructed to spell each syllable letter by letter or to read each number digit by digit as they appeared in the window.

A list of three syllables and three numbers was used to practice the procedure, after which the experimental list was presented.

Materials.—The stimulus lists consisted of ten syllables and ten three-digit numbers and are shown in Table 1. There are two series, A and B. All syllable-number positions of Series A are reversed in Series B so that the effects of crowding and isolation can be determined independently of the type of item. Series A was used with half the *S*s, Series B with the other half. According to the perceptual analogy, Items 3 and 13 are isolated items embedded in a homogeneous series.

The items were presented on a Hull-type memory drum at the rate of one every 4 sec. This rate of exposure was required to allow sufficient time for the reading of the item by *E* and repetition by *S*. The list was presented only once.

TABLE 1

The Stimulus Lists

Serial Position	Series A	Series B	Serial Position	Series A	Series B
1	GUB	341	11	341	GUB
2	KEV	258	12	258	KEV
3*	406	SIH	13*	SIH	406
4	DAC	179	14	179	DAC
5	RUL	417	15	417	RUL
6	HOF	562	16	562	HOF
7	763	FIP	17	FIP	763
8	VOM	738	18	738	VOM
9	581	TER	19	TER	581
10	WAJ	269	20	269	WAJ

* Relatively isolated items.

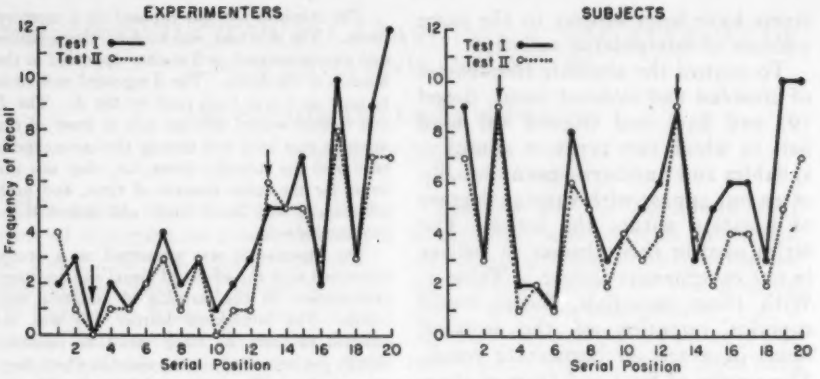


FIG. 1. Number of correct reproductions by *Es* (incidental learners) and *Ss* (intentional learners) at each serial position in Exp. I. Isolated items are indicated by arrows.

Retention tests.—Immediately after the presentation of the stimulus list, the first retention test was given. Both *S* and *E* were required to write down as many of the syllables and numbers as they could remember. They were requested to write down the items regardless of order of presentation and urged to put down guesses as well as items of which they felt certain. There was a time limit of 5 min.

Following the first retention test, both *Ss* and *Es* were engaged in an interpolated task. This task was presented as a test of reading speed and comprehension. They were handed marked passages from Plato's *Dialogues*, and at a sign from the investigator started reading at their normal speed.

After 20 min. of reading, a second retention test was given. The *Ss* and *Es* were again required to write down as many syllables and numbers as they could remember. The time limit was again 5 min.

Postexperimental inquiry.—At the end of the experimental session, all *Ss* and *Es* were questioned concerning their interpretation of the experimental situation. A standard set of questions was used. The *Es* were asked whether or not (*a*) they had expected the first recall test; (*b*) they had attempted to memorize any of the items during the presentation of the series; (*c*) there had been any items that had caught their attention during the presentation of the series; (*d*) they had expected the second retention test; (*e*) they had attempted to rehearse any of the items between the first and second retention tests. The *Ss* were asked only questions (*c*), (*d*), and (*e*). On the basis of this interview, *Es* who had anticipated the first recall test (about 20%) were eliminated from the group of incidental learners, and were replaced.

Subjects.—There were 28 pairs of *Ss* and *Es*. All were undergrauate students at the University of California.

Results

Amount of retention.—Figure 1 presents the number of *Ss* and *Es* who correctly recalled each of the items on the first and second retention tests. Table 2 shows the mean number of items recalled by *Ss* and *Es* on the two tests.

On the first retention test, which may be considered a measure of original learning, *Ss* recalled significantly more items than *Es* ($t = 2.50$, $p = .01$). Both groups show a significant drop in recall on the second test. Tests of significance of the

TABLE 2

NUMBER OF ITEMS CORRECTLY RECALLED BY *Ss* AND *Es* IN EXP. I

Group	Test I		Test II	
	Mean	*SD*	Mean	*SD*
Es (Incidental)	2.78	1.47	2.21	1.29
Ss (Intentional)	3.89	1.76	3.00	1.88

retention loss yield a t of 3.31 for *Ss* ($p < .01$) and a t of 2.19 for *Es* ($p = .03$).

When the level of original learning is taken into account, there is no significant difference between incidental and intentional learners in the amount of retention loss. Analysis of covariance of the scores on the second retention test yields an F ratio of .01; i.e., the difference in retention loss between the two groups is entirely accounted for in terms of the difference on the test of original learning.

Serial position effects.—Inspection of Fig. 1 suggests important differences between the serial position curves for incidental and intentional learners. The curve for *Ss* shows pronounced rises in the frequency of recall both at the beginning and the end of the list and in that respect resembles the typical bow-shaped serial position curve. The curve for *Es*, on the other hand, shows no primacy effect whatsoever but a very pronounced recency effect. Although the over-all level of retention is lower for *Es*, their curve surpasses that for *Ss* in the final position of the series. It appears that the primacy effect depends on intention to learn, whereas the recency effect does not.

The effects of isolation.—We now turn to the effects of relative isolation on retention under conditions of intentional and incidental learning. Reference to Fig. 1 suggests at once that such effects, if any, occur *only under conditions of intentional learning.*

Consider intentional learners first. The isolated items, 3 and 13, represent peaks in the retention curve of *Ss*. On the first test, the frequency of recalls at both isolated positions is 9, or 32.14%, of the *Ss*. Only items at the very beginning and end of the series equal or exceed this level of recall. A control experiment by Siegel (**9**) had shown that neither the particular items used nor their serial positions per se can account for the peaks in the isolated positions. The superior retention of these items must, therefore, be attributed to their relative isolation in the series. It should be noted that the peak in Position 7 is also for a relatively isolated item.

In finding superior immediate retention for both isolated items, we agree with the findings of Siegel and only partly with those of Saul and Osgood (**8**). The latter investigators report a peak at Position 3 but not at Position 13. To test the significance of the peaks in Positions 3 and 13, the percentages of recall at these positions were compared with the percentages of recall for items that immediately precede and follow them. These comparisons hold the effects of serial position as constant as possible. The test of significance for differences between correlated proportions was used (**7**, pp. 77 ff.). Recall for Item 3 is significantly higher than recall for Item 2 ($CR = 2.46$, $p = .01$) and Item 4 ($CR = 2.34$, $p = .02$). The difference between Items 12 and 13 is not significant ($CR = .83$, $p = .40$), but recall is significantly higher for Item 13 than for Item 14 ($CR = 2.34$, $p = .02$).

There is some slight evidence for superior retention of the isolated items on the test of delayed recall. As Fig. 1 shows, there was no loss in retention for either Item 3 or Item 13, whereas items with a comparable degree of original retention, viz., those at the beginning and end of the list, show substantial drops in retention. There are, however, several other items that show no retention loss at all, although comparison with the isolated items is made difficult by the variations in degree of original retention.

The picture is altogether different for incidental learners (*E*s). As Fig. 1 clearly shows, there is no indication whatsoever of superior recall for the isolated items on the first test. Not a single incidental learner recalled Item 3 (which is, indeed, the only item for which there is zero recall), nor is there a peak in Position 13. The recalls for Items 13 and 14 are, in fact, exactly equal, nor is there a significant difference between Items 12 and 13 ($CR = 1.42$, $p = .16$).[1] The second retention test does not change the picture. Retention for Item 3 is still zero. There is a gain of one in the recall score for Item 13, but there is another nonisolated item (No. 1) that shows such a reminiscence effect, and there are several items that show no retention loss. Clearly, then, isolation has significant effects under conditions of intentional learning, but not under conditions of incidental learning.

Postexperimental inquiry.—In addition to identifying those incidental learners who expected the first retention test, the postexperimental inquiry yielded the following information: Only 1 *E* and 3 *S*s out of 28 pairs stated that they expected the second retention test. Four *E*s and 5 *S*s admitted to some rehearsal between the first and second retention test. The retention losses of these individuals are, however, comparable to those of other members of their groups.

In answering the question about items that especially caught *S*s' and *E*s' attention, not a single individual mentioned, directly or indirectly, the fact of perceptual isolation! The large majority of both *S*s and *E*s did mention at least one item that had caught their attention during the presentation of the list. In virtually every case, the reason for which an item was especially noticed was similarity to a meaningful word or symbol. Some of these associations referred to conventional meanings, e.g., English words; other associations were idiosyncratic, such as initials or house numbers. In any event, as far as the reports are concerned, associative context rather than perceptual isolation determined selective attention to specific items in the series.

The problem of degree of isolation.—This experiment has failed to obtain significant effects of crowding and isolation under conditions of incidental learning, i.e., under the very conditions which gestalt theory would regard as most favorable to "spontaneous organization." More precisely, no such effects were found after one reading of the list, i.e., at an early stage of learning. A single training trial was, however, sufficient to produce significant differences between crowded and isolated items under conditions of intentional learning.

The question still may be raised whether a sufficient degree of isolation was provided. In reply to such a criticism it might be pointed out, of course, that positive results obtained with the stimuli used in this study have been accepted as evidence for the effects of isolation (**9**). Negative results obtained with the same stimuli under conditions of incidental learning

[1] Further tests of significance were carried out, comparing the differences in the effects of isolation between groups. The difference in the proportions of recall between Positions 2 and 3 is significantly greater for intentional than for incidental learners ($CR = 2.46$, $p = .02$) as is the difference between Positions 3 and 4 ($CR = 2.34$, $p = .02$). The groups do not differ significantly when the differences between Positions 12 and 13 are compared ($CR = .22$). In the case of the drop between Positions 13 and 14, the difference in favor of the intentional learners yields a CR of 1.70 ($p = .09$).

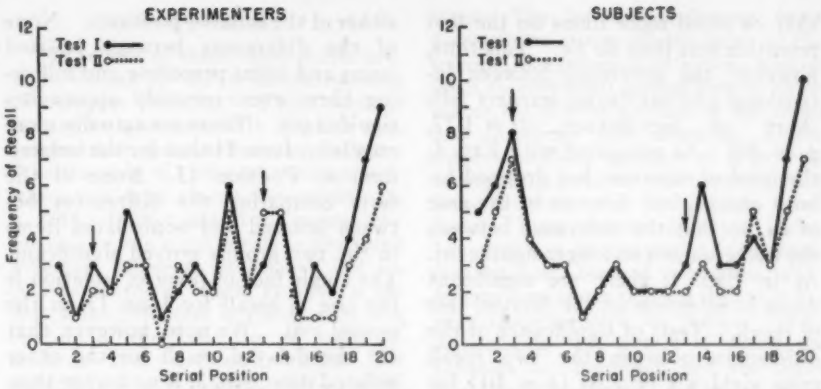


FIG. 2. Number of correct reproductions by *Es* (incidental learners) and *Ss* (intentional learners) at each serial position in Exp. II. Isolated items are indicated by arrows.

are at least equally valid. Nevertheless, it seemed advisable to repeat the experiment with stimulus materials providing a higher degree of perceptual isolation than those used in the present experiment. The following section reports the results of this further study.

EXPERIMENT II

Method

Materials.—The stimulus lists were the same as in Exp. I, with the additional feature that in each series one type of item was printed in red, and the other in black. For half the *Ss* and *Es* learning a given series (A or B), the nonsense syllables were printed in black and the numbers in red; for the other half, the colors were reversed. Isolated items now not only belonged to a different class than the surrounding ones but also differed from them in color.

Procedure.—The procedure was in all respects identical with that used in Exp. I. One question was added to the standard interview given during the postexperimental inquiry. At the end of the interview both *Ss* and *Es* were questioned concerning their perception of the color of the items.

Subjects.—There were 28 pairs of *Ss* and *Es*. All were undergraduate students at the University of California.

Results

Amount retained.—Figure 2 shows the number of *Ss* and *Es* correctly recalling each of the items on the two tests of retention. Table 3 presents the mean number of items recalled on the two tests.

The general level of retention is lower in Exp. II than in Exp. I for both *Ss* and *Es*. When the scores on the first retention test are compared, *Ss* in Exp. II are significantly lower than *Ss* in Exp. I ($t = 2.45$, $p = .02$). The difference is not significant in the case of *Es* ($t = 1.03$, $p = .30$). The introduction of the color difference reduces the amount of recall for all *Ss* and *Es*. It is plausible to conclude that the variation in color served as a source of intraserial interference and did not favor differentiation of the stimulus items.

Turning now to a comparison of the groups in Exp. II, we again find

TABLE 3

NUMBER OF ITEMS CORRECTLY RECALLED BY *Ss* AND *Es* IN EXP. II

Group	Test I		Test II	
	Mean	*SD*	Mean	*SD*
Es (Incidental)	2.39	1.42	1.89	1.37
Ss (Intentional)	2.86	1.33	2.36	1.14

that *S*s recall more items on the first retention test than do *E*s. This time, however, the difference between intentional and incidental learners falls short of significance ($t = 1.27$, $p = .20$). As compared with Exp. I, the level of retention has dropped for both groups, but more so in the case of *S*s, so that the difference between the two groups is no longer significant. As in Exp. I, there are significant drops in retention on the delayed test of recall. Tests of significance of the differences between the two recall tests yield a t of 2.50 ($p = .01$) for *S*s and a t of 2.63 ($p < .01$) for *E*s. Analysis of covariance of the scores on the second recall test shows that the two groups do not differ in the amount of forgetting when the level of original learning is taken into account ($F = .03$).

Serial position effects.—Except for the generally lower level of recall, the serial position curves have the same general form as in Exp. I. Again we note that the curve of *E*s shows a pronounced rise at the end but not at the beginning of the series.

Effects of isolation.—Inspection of Fig. 3 shows that the introduction of the color difference has not substantially increased the effects of isolation for incidental learners and has actually reduced such effects for the intentional learners.

On the first retention test for *S*s there still is a peak in Position 3, but the percentage of recall at that position is not significantly higher than the percentage of recall at Position 2 ($CR = .58$, $p = .50$) or Position 4 ($CR = 1.42$, $p = .16$). The peak at Position 13 has disappeared entirely. There is no evidence for superior recall of the isolated items on the second retention test.

The first retention test of the *E*s fails to produce a significant peak at either of the isolated positions. None of the differences between isolated items and items preceding and following them even remotely approaches significance. There are actually more recalls for Item 14 than for the isolated item at Position 13. None of the tests comparing the differences between isolated and nonisolated items in the two groups proved significant. The single fact in favor of isolation is the rise in recall for Item 13 on the second test. We note, however, that (*a*) the delayed recall for the other isolated item, No. 3, is no higher than for crowded items with comparable scores on the first test, and (*b*) there is reminiscence for another item (No. 8), and several items show no loss in retention. The net evidence does not point to isolation as a significant determiner of incidental memory.

Postexperimental inquiry.—The results of the postexperimental inquiry are similar to those in Exp. I. None of the *E*s and seven of the *S*s stated that they expected a second retention test. One *E* and four *S*s admitted to some rehearsal between the two tests, but their delayed retention scores are comparable to those of other members of their groups. In commenting on items that especially caught their attention during the original presentation, no *S* or *E* made direct or indirect reference to perceptual isolation. All the items mentioned were singled out for special attention because of their similarity to meaningful words and symbols.

All but four *E*s and two *S*s recalled the fact that the syllables and numbers differed in color. Accuracy in identifying the colors of the two classes of items was, however, surprisingly low. Only nine *E*s and nine *S*s could state positively which of the items were red and which were black.

Discussion

The results of these experiments present evidence for the incidental learning of nonsense material.[2] These findings are not in agreement with Köhler's suggestion that without intention to learn "the learning of series of nonsense syllables would be almost impossible" (4, p. 285). In fact, one of the noteworthy findings of the experiments is the small difference between intentional and incidental learners. Such, at least, is the case when retention is tested after one reading of the list. Use of only one training trial necessarily limits the generality of our findings. This restriction was imposed in the belief that the shorter the training period, the less likely were the incidental learners to develop a covert set to learn. An early test may, therefore, be maximally sensitive to the difference between intentional and incidental learning.

Our results do not agree with the view that factors favoring perceptual organization, such as the isolation of an item in a homogeneous series, are necessarily more effective under conditions of incidental than intentional learning. In fact, we find quite the opposite. It was only under conditions of intentional learning that any significant effects of isolation were obtained. Such effects were, moreover, largely limited to the case in which isolation was only a matter of the serial order of two classes of items (Exp. I). Addition of a color difference reduced the effects of isolation (Exp. II).

We suggest that the relative isolation of an item in a stimulus series favors retention only to the extent that the stimulus features producing isolation are relevant to the learner's task. If isolation is along dimensions irrelevant to the learner's task, its effects will be negligible or even negative.

As Gibson (1) has pointed out, the learning of a series of stimuli requires discrimination among the individual items. The intentional learner must discriminate among all the items of the series. The presence of similar stimuli leads to intraserial generalization which lowers the discriminability of the items and interferes with learning. Under these conditions isolation of a syllable among numbers, or vice versa, should favor discrimination by reducing interference from closely adjacent items. Isolation aids discrimination, and therefore learning, because it reduces interference from items that are similar in the very characteristics the learner has to discriminate—the letters of which the syllables are composed and the digits forming the numbers. By this analysis, the separation of black items from red ones should not increase retention because discrimination of this color difference is not relevant to the performance of the learning task. In fact, the color difference is a source of interference because it leads the learner to respond to stimulus features that are irrelevant to the learning task. Generalization along this irrelevant dimension adds to the amount of intraserial interference.

This analysis, in terms of associationistic principles, accounts for the beneficial effects of isolation for intentional learners in Exp. I and also for the reduction in these effects in Exp. II. The fact that isolation has little or no beneficial effects for incidental learners is also readily understandable. The incidental learner does not need to, and we presume does not attempt to, make as many differential responses to the individual items as does the intentional learner. Hence, the incidental learner forms not only fewer correct associations than the intentional learner but also fewer incorrect ones. There is less intraserial interference, and therefore less reduction in such interference by the isolation of an item. It appears that the incidental learner responds primarily to the associative context aroused by individual items, such as similarity to meaningful words and symbols, and not to the differences among items. The results of the postexperimental inquiry bear out this conclusion.

[2] Many of the early experiments used meaningful materials. Jenkins (2) used nonsense syllables.

Our results are clearly at variance with the gestalt hypothesis that the formation and development of memory traces are governed by laws of perceptual organization. Whatever effects of perceptual isolation were found can be adequately accounted for in terms of the principles of associative interference.

Summary

Experiments were performed to compare memory for "crowded" and "isolated" materials under conditions of intentional and incidental learning. In Exp. I, *Ss* learned lists of ten syllables and ten numbers. Isolation was achieved by embedding a number in a homogeneous sequence of syllables, and vice versa. Isolated items were retained better than crowded ones by intentional learners but not by incidental learners. In Exp. II, the same stimulus materials were used but a color difference was added to increase the degree of isolation of the critical items. Under these conditions there were no significant effects of isolation on either intentional or incidental memory.

It is concluded that isolation favors retention only to the extent that the stimulus features producing isolation are relevant to the learner's task. The limited effects of isolation that are found can be adequately accounted for in terms of intraserial interference and do not support the gestalt theory of memory.

References

1. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *J. exp. Psychol.*, 1940, **47**, 196–220.
2. Jenkins, J. G. Instruction as a factor in "incidental" learning. *Amer. J. Psychol.*, 1933, **45**, 471–477.
3. Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
4. Köhler, W. *Gestalt psychology*. New York: Liveright, 1929.
5. Köhler, W., & Von Restorff, H. Analyse von Vorgängen im Spurenfeld: II. Zur Theorie der Reproduktion. *Psychol. Forsch.*, 1935, **21**, 56–112.
6. McGeoch, J. A., & Irion, A. L. *The psychology of human learning*. New York: Longmans, Green, 1950.
7. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
8. Saul, E. V., & Osgood, C. E. Perceptual organization of materials as a factor influencing ease of learning and degree of retention. *J. exp. Psychol.*, 1950, **40**, 372–379.
9. Siegel, P. S. Structure effects within a memory series. *J. exp. Psychol.*, 1943, **33**, 311–316.
10. Von Restorff, H. Über die Wirkung von Bereichsbildung im Spurenfeld. Analyse von Vorgängen im Spurenfeld. I. Von W. Köhler und H. v. Restorff. *Psychol. Forsch.*, 1933, **18**, 299–342.

(Received October 28, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

MAXIMUM RATES OF FORM PERCEPTION AND THE ALPHA RHYTHM: AN INVESTIGATION AND TEST OF CURRENT NERVE NET THEORY

ODDIST D. MURPHREE[1]

Veterans Administration Hospital, North Little Rock, Arkansas

Wiener (7) emphasizes the importance of synchronized scanning processes in the new science of cybernetics. He is also inclined, along with Culbertson (2) and McCulloch and Pitts (5), to view the human brain as essentially an extremely complex computing or "information" machine. It becomes of paramount interest, then, to inquire about synchronous neurologic processes that the brain possesses. In electronic or other type computing machines, one of the main purposes served by the synchronous scanning processes is to reduce, by a large factor, the amount of material and space comprising the machine. For this reason it is plausible to expect widespread use of synchronous processes in human brains since their relatively small size makes them the most efficient information machines yet known (7, p. 55 ff.).

The cyberneticists mentioned above consider the alpha rhythm (basic resting occipital rhythm) to be one of the chief neurologic synchronous processes, and the perception of shape or form to be its most important area of function. This study was designed to subject to experimental test certain hypotheses regarding the part the alpha rhythm is supposed to play in form perception. The following are the two hypotheses relevant in the present context:

1. The alpha rhythm is an index of certain neurologic processes necessary for the transmission of neural patterns that mediate the perception of form. These processes are periodic in nature and, when measured in terms of electrical potential, constitute the alpha rhythm. These processes are the basis for the loss in temporal acuity in rapid successive form perceptions.

2. Apparent motion is a manifestation of a partial failure or loss in efficiency in the perceptual process by the retinocortical nerve net in the transmission of spatial and temporal information. There is an interactional equivalence of the space and time information (in neural terms) such that neither of these is presented accurately for perception, and the neural pattern for the perception of motion results.

From these hypotheses, in conjunction with certain assumed characteristics of the system, the following may be deduced:

1. Separate form images may not be perceived at a rate faster than the alpha frequency for a particular individual.

2. Differing forms presented successively at rates faster than alpha frequency will be perceived as in motion or in simultaneous existence, depending on the ratio of their rate of presentation to the alpha frequency of a particular individual.

3. That rate of form presentation at which apparent motion ceases (threshold

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his study and do not necessarily reflect the opinion or policy of the Veterans Administration.

This study is based on a dissertation presented to the University of Cincinnati in partial fulfillment of the requirements for the Ph.D. degree.

The author is indebted to Dr. Ranald M. Wolfe, Veterans Hospital, Chillicothe, Ohio, for his cooperation in making the investigation possible.

of simultaneity) will be proportionate to the alpha frequency.

4. The maximum rate at which separate images can be perceived will be proportionate to the alpha frequency.

5. Segments, of a composite form, which may combine to make up a unit form, when presented tachistoscopically at or above alpha frequency, will be seen as fused into a composite form whose elements are simultaneously present.

Method

The experiment was designed to test each of the above corollaries. In addition, since the opportunity was easily available, the design included a figure-ground reversal procedure. It was hoped to determine whether or not the alpha frequency was in some meaningful way related to the frequency of reversal of a figure and ground that was sufficiently rapid for the simultaneous perception of a blended or fused area. Critical flicker fusion (CFF) of "on-off" light on a similar area was also obtained in order to rule out "retinal fusion" and other nonspatial interference factors. The fusion of the forms should take place well below CFF.

Subjects.—Fifty *Ss* were selected from the names on file with the electroencephalographer at the Veterans Administration Hospital, Chillicothe, Ohio, on the basis of *Ss'* alpha frequency as determined by the neurologic consultant. It is extremely difficult to obtain *Ss* who possess frequencies below 8 or above 12 per sec. and and who do not also have a diagnosis of organic pathology. The distribution obtained exhausted the extreme ranges from the more than 2,000 tracings on record at the hospital. Table 1 gives the number of *Ss* studied in each alpha frequency grouping.

It is assumed that the significant differences between the groups are limited to their alpha frequencies or that those variations which were present influenced the study in no systematic way. The *E* deliberately remained unaware of the alpha frequency of any *S* until after the data were collected, because he worked rather closely with *Ss* in the operation of the apparatus and could have been instrumental in modifying their responses. The groups were completed in a random fashion so that the effects of practice and other variables would be distributed to all groups.

TABLE 1

Number of *Ss* in Relation to Alpha Frequency

Frequency	No. of *Ss*
12–13	4
11–12	12
10–11	9
9–10	13
8–9	12
Total	50

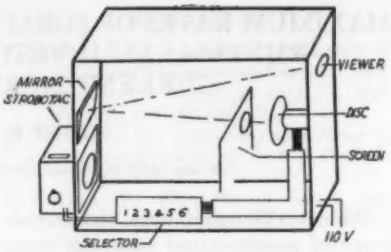


Fig. 1. Apparatus in near light-tight box. Distance from viewer to disc through mirror is 94 cm.

Apparatus.—The apparatus was essentially a device which could be made to expose and illuminate various stimuli at controlled and variable rates (see Fig. 1). The heart of the apparatus consisted of a spinning disc onto which the flash of a General Radio Strobotac (Type 613-B) was directed through the window of a screen. The strobotac emitted a 5-msec. flash, either at a determined frequency or whenever an external contact was operated. An electric commutator on the same axis with the spinning disc was the external contactor in this case. An array of switches enabled *E* to select the particular stimuli to be exposed. In operation, when the desired stimulus appeared in the window of the screen, it was automatically illuminated in a manner quite analogous to the firing of appropriate cylinders in an internal combustion engine. It should be pointed out that the strobotac also operated as a tachometer, and by this means the speed of rotation of the disc was quickly determined.

Procedure.—A standard procedure was completed by each *S* individually.[2] The *S* entered the room; about 2 min. of general conversation and explanation ensued, chiefly to put *S* at ease and insure cooperation; and his eyes were adapted to the subdued illumination. The *S* then placed his head at the "viewer" of the apparatus and the various thresholds were obtained in the following manner: While *S* fixated the white

[2] A more detailed description of the procedures is presented in the original dissertation on file at the library of the University of Cincinnati.

area in the window, he was required to say "now" when the illumination (from the strobotac) ceased flickering and again when it started flickering. The average between the two was considered his CFF frequency. In like manner, his rates were recorded where apparent motion ceased and the stimuli (two white circular areas alternately appearing in adjacent positions) seemed simultaneously present. That is, he was required to state when the motion stopped and when it began again. Next, *S* watched a rapidly writhing, wiggling blob or lighted area, which was the result of separate forms being successively exposed and illuminated, though too rapidly for recognition. There were four separate forms, and one followed the other in an unbroken sequence. The forms were simple: a hexagon, circle, half-circle, and a "football" shape; they were approximately equal in area. The *S* was asked to identify the shapes (or tell what he saw) as the exposure rate was gradually reduced until all four forms could be identified. Appropriate rates were recorded. After this, while *S* watched, a black circle on a white circular background was exposed alternately with a white circle on a black circular background. These alternating stimuli were congruent. As the exposures became more rapid, the distinct percepts of figures on grounds gave way to the perception of a blended or fused area. The threshold was crossed going up and down. In the last procedure *S* indicated when successively presented segments of a composite form were exposed sufficiently rapidly for the composite form to be perceived as a simultaneous mosaic, and, again, when the exposure rate was sufficiently slowed for the mosaic to break up into temporally discrete parts.

Results

The relation between the alpha frequency and the various performance measurements tends to be similar for all the procedures except CFF. The relationship is apparently curvilinear. The correlation ratios of form recognition and apparent motion rates with the alpha rhythm were significant at the 1% level; for the figure-ground reversal with alpha, the confidence level is between the 1% and 5% level; for the fusion rates of the various elements into composite forms with alpha frequency, the confidence level was not significant at the 5% level.

The over-all picture is seen best in Fig. 2. The means of the performances of the different groups for thresholds of form identification, apparent motion, figure-ground reversals, and fusions (of two element composite form) are plotted. Each line represents the mean performance of all *S*s on a particular procedure. The group displaying 12–13 per second alpha frequencies is composed of only four members, and though it may be less reliable, fits the trend of the data. Subjects in this frequency range are rare.

While all the procedures did not pass rigorous tests of significance, they display the same general relation, with the exception of CFF. This phenomenon took place at such high frequencies, generally higher than 40 per second, that there is no possibility of assigning this type of fusion to the same source as the others. It is seen that these others stand in a specific relation to the alpha frequency.

A curve of best fit is reasonably

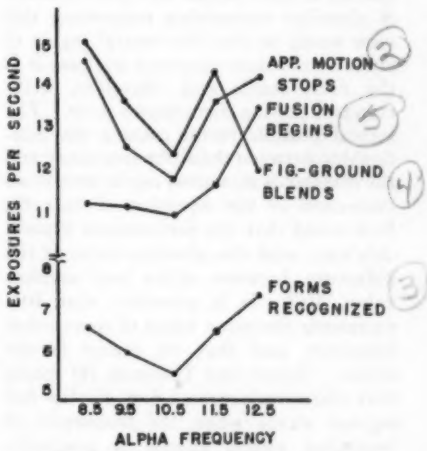


Fig. 2. Relation between alpha rhythm and performance

approached by equations of the second order, though the best result is apparently had by the first-order equation

$$P = K + |\text{alpha} - 10.3|$$

where P equals performance or events per second at which perceptual thresholds occurred, and K equals a constant which has a particular value for each type of procedure. The vertical bars signify that the absolute rather than the algebraic value is to be considered. It is of theoretical interest that the constant 10.3 is this value for all performances.

Discussion

The general relation found between the perceptual thresholds and the alpha frequency might result from the operation of either of two factors. First, the curves (Fig. 2) may be depressed in the center by the central range of alpha. This frequency is by far the most common in the population and may be different in quality. A statement by Gibbs (3) is perhaps relevant. It is in effect an assertion that in a system of oscillators with equal amplitudes, that oscillator whose frequency is most stable will predominate. A plausible assumption supporting this view would be that the central region of alpha is the most common because it is the most stable and, therefore, withstands flickering disturbance most. The second possible factor causing the relationship between the alpha frequency and the thresholds measured can be seen in an inspection of the equation of best fit. It is noted that the performance thresholds vary with the absolute value of the difference between alpha and another value, 10.3. It is plausible that 10.3 represents the mean value of some other frequency and that its source is the retina. Granit and Therman (4) found that electroretinograms first display full regular waves when the frequency of impinging photic stimuli is gradually reduced to 10.3 per second. Bartley's statement that "it has become evident that the retina possesses an intrinsic rhythm . . . [with] sensory significance, . . ." (1, p. 328) supports the above assumption. It appears that the span of simultaneity, or maximum temporal separation of two events that are perceived as simultaneous, is lengthened for those *S*s whose alpha and retinal frequencies approach the same value. Therefore, with the span of simultaneity possibly dependent on both the alpha and retinal frequencies, the next step would be to find out more about the interactional results of the two frequencies. Certainly a modification in theory as presented in the first part of this paper is indicated. However, the mean span of simultaneity for all 50 *S*s was found to be .095 sec. The mean span of their alpha cycles is .098 sec. This is strongly suggestive that the alpha rhythm is, in part, responsible for this quantum factor in rapid successive perceptions.

Since retinal rhythms often are of the same order and near alpha rhythms in frequency, it might be asked if the retinal rhythm might not be primary in determining the quantum factor in the above perception. This question is answered, in part, by the significant correlation between the frequency of the stimuli at perceptual thresholds and the alpha frequencies. We do not know if the retinal rhythms would also correlate highly with alpha.

The hypotheses in this study were taken from the field of cybernetics. This emphasizes a point often made by cyberneticists themselves that theirs is a broad field with wide implications for other sciences. A productive and effective relationship extremely profitable for psychology seems very probable. Those readers desiring a philosophical treatment of cybernetics are referred to Northrop's (6) evaluation of McCulloch and Pitts's work.

Summary

Cyberneticists and mathematical biophysicists have hypothesized synchronous scanning neurologic processes, particularly the alpha rhythm, which mediate the visual perception of form or

shape. From these hypotheses certain deductions (for the most part concerned with a quantum time factor as implied in any scanning process) were made and experimentally tested.

The results show:

1. There is a significant relation between the alpha rhythm and the temporal aspect of rapid successive spatial perceptions. This relation may be considered as curvilinear, or, possibly, linear to the difference between alpha and some other value, which in central tendency nearly coincides with alpha.

2. Strong support is indicated for the view that maximum rates of form perception, apparent motion, figure-ground reversals (at appropriate rates), and fusion of separate spatial elements into simultaneously perceived composite forms or shapes are mediated by a nerve net of which the alpha rhythm is an integral part.

References

1. Bartley, S. H. *Vision: a study of its basis.* New York: Van Nostrand, 1941.

2. Culbertson, J. F. *Consciousness and behavior: a neural analysis of behavior and consciousness.* Dubuque, Iowa: Wm. C. Brown, 1950.

3. Gibbs, F. A., & Gibbs, E. L. *Atlas of electroencephalography.* Cambridge, Mass.: Addison-Wesley Press, 1950. Pp. 22–25.

4. Granit, R., & Therman, P. O. Excitation and inhibition in the off-effect of the retina. *J. Physiol.*, 1937, **91**, 127–139.

5. McCulloch, W. S., & Pitts, W. How we know universals: the perception of auditory and visual forms. *Bull. math. Biophysics*, 1943, **5**, 115–133.

6. Northrop, F. S. C. The neurological and behavioristic psychological bias of the ordering of society by means of ideas. *Science*, 1948, **107**, 411–416.

7. Wiener, N. *Cybernetics.* New York: Wiley, 1948.

(Received October 8, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

AN ATTEMPT TO OBTAIN PUPILLARY CONDITIONING WITH INFRARED PHOTOGRAPHY[1]

FRANCIS A. YOUNG

The State College of Washington

In the controversy that has developed over the question of whether or not the contraction or dilation of the human pupil can be conditioned to a nonadequate stimulus, only three investigators (**1, 2, 5**) have consistently found pupillary conditioning to a sound occurring in most or all the subjects employed. On the other hand, seven investigators (**3, 4, 8, 9, 10, 11, 14**) have failed to support these results. Of the three experiments in which conditioning took place, the two by Cason (**2**) and by Hudgins (**5**) used similar techniques, while the one by Baker (**1**) utilized an entirely different methodology and reported conditioning taking place in only two or three trials in contrast to the several hundred trials found necessary by Cason and Hudgins. The present study is based upon the findings of Cason and Hudgins, with certain modifications in apparatus and technique. An excellent summary of the experiments performed with respect to pupillary conditioning may be found in the study by Stern (**10**).

In all the studies indicated above, the measurements of the pupil were made with some type of pupillometer, which required *E* to keep two hair lines in contact with diametrically opposite edges of the pupil or with the center and one edge of the pupil. As Steckle and Renshaw (**8**) have pointed out, this is an extremely difficult task which requires the utmost concentration on the part of *E*. The inability of *E* to maintain this level of concentration, and the resultant variability of *E* and his measurements, are undesirable factors in this type of experiment. In the present study these variables have been eliminated by photographing the pupil during the test trials and taking the measurements from the photographs.

The experimental conditions employed by Cason were not clearly stated in his paper, an omission which makes repetition and interpretation difficult. The conclusions which he drew from his investigation are: (*a*) The pupillary reflex can be conditioned to the sound of an electric bell, to the buzzing sound of a telephone receiver, or to an electric shock. (*b*) A contraction of the pupil is more definitely conditioned than a dilation. (*c*) The CS may eventually cause a change in the size of the pupil that can be readily distinguished by the observer. With the procedure used, this change generally amounted to about .5 mm. (in horizontal diameter); but it differed widely in different *S*s and in the same *S* at different times. (*d*) A CR is readily established when the bell stimulus is given during a change in the size of the pupil. Every evidence is found in favor of this conclusion. No very good evidence is found in favor of the supposition that a CR may be established when the bell stimulus is given before the change in the size of the pupil.

Although Cason conditioned both pupillary contraction and dilation, the present study will be concerned only with contraction in agreement with conclusion *b* above, and will follow the dictum of conclusion *d* with respect to the pairing of the UCS and CS. The salient features of Cason's procedure may be summarized as follows: the use of massed practice, lack

[1] This investigation was supported in part by funds provided for biological and medical research by the State of Washington, Initiative measure no. 171.

of a fixation point to keep convergence and accommodation constant, the same light for measuring that was used in training, no specification of brightness of the light source in the training sessions, measurements taken before the training trials and at the end of 400 trials, and a light-to-darkness sequence.

Hudgins modified Cason's procedure by using two levels of light intensity, a training light of 37.7 meter-candles and a test light of 3.5 meter-candles.[1] This light arrangement permits measurement of the partially dilated pupil with a telescopic measuring device similar to that used by Cason. Hudgins conditioned the pupil to a variety of stimuli other than a sound; these variations were not repeated in this study. The conditioning to the bell employed as a sound stimulus by Hudgins was accomplished in two experimental periods extending over two days. During the first period which lasted 1.5 to 2 hr. *S* was given 125 training trials. A typical training trial was as follows: the light and bell circuits were closed simultaneously by *E*. The stimuli remained on for a period of 3 to 4 sec. This was followed by a period of 10 sec. of darkness and silence. At the end of the group of 125 trials several test measurements were made. On the second day 80 to 100 additional trials were given, and the final tests were made at the end of these trials.

The experimental procedures of Cason (2) and of Hudgins (5) are similar in that both used a light-to-darkness sequence, no fixation point during the period of darkness, a pupillometer for measuring the size of the pupil, and massed trials, and both took measurements before and at the end of the training period. Hudgins' description of the conditioned contraction reveals that it has a latency of 2 sec., which is about ten times that usually found for the light reflex. A second fact which appears is the relatively slow time of contraction, with an average of 10.93 sec. and an *SD* of 2.75 sec. Again, since we are primarily interested in pupillary contraction, Hudgins' findings with respect to pupil dilation have been ignored.

Apparatus and Procedure

An infrared camera designed to operate automatically at any speed up to 20 exposures per second was used in this study. This apparatus is described in detail elsewhere (12). The *S* was seated on an adjustable chair with his head in a head rest and a bite board between his teeth. A small millimeter rule carrying a subject identification symbol was located just below the lower eyelid in the same vertical plane as the pupil. This ruler served as a guide for size since the pictures were later enlarged 20 times by means of a film strip projector for measurement of pupillary diameter. The camera was mounted in the same horizontal plane as the pupil of *S*'s eye. Immediately below the camera lens and 3° below the horizontal was mounted a small rectangular mirror (1 × 2 in.). By means of this mirror, *S* continuously fixated, during both light and dark phases, a small red bull's-eye consisting of a center dot and three equally spaced concentric circles, the largest being ¾ in. in diameter This fixation target was mounted above and behind *S* at a total distance of 1 m. from the eye, and was kept as dim as possible yet still visible when the UCS light was on. The light which served as the UCS was mounted directly below the mirror and camera lens at an angle of 20° with the horizontal. The light itself was a circular piece of ground glass transilluminated by a 60-w. frosted bulb enclosed in a light-tight box. The luminance of the light was 4,871.0 millilamberts.

All preliminary and conditioning tests were made with the camera set for three pictures per second. Ten male and ten female *S*s, all between 20 and 33 years old, were given 400 training trials in eight units of 50 trials. At least 1 hr., and not more than 24 hr., elapsed between each set of trials. A test of conditioning was made at the end of every set in order to determine when and in what manner conditioning occurs. In this way the massed practice employed by both Cason and Hudgins is retained without encountering the fatigue effects described by Lowenstein (6). Lowenstein found that the pupil shows a reduction in the magnitude of response to a light-darkness sequence of 1-sec. light and 3-sec. darkness within a relatively few trials, and reaches a state of complete immobility after about 60 repetitions of this sequence. To reduce still further any possibility of fatigue modifying the results, a light-darkness sequence of 4-sec. light and 25-sec. darkness was used in this study.

The CS was a 1,024-cps tone produced by a 6BP Western Electric audiometer at an intensity level 60 db above *S*'s threshold. The limen for each *S* was determined with the same audiometer using the method of limits. The onset of the

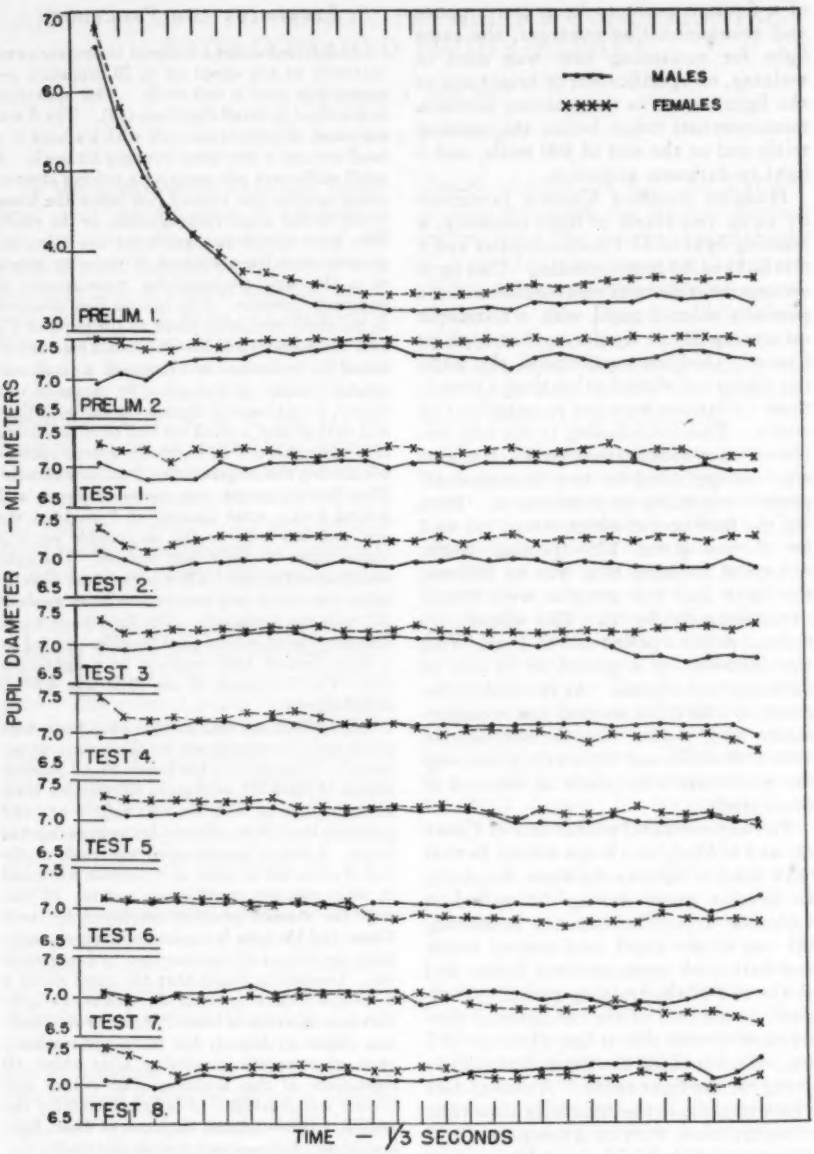


FIG. 1. Preliminary Test 1 was made with the UCS (white light) but with no CS (tone) present. Preliminary Test 2 was made with the CS (tone) alone. Tests 1 through 8 were made at the end of the respective sets of 50 training trials.

tone, which was presented by means of earphones, preceded the onset of the light by 1 sec.; the tone continued and stopped at the same time that the light was turned off. Thus the total cycle of tone—5-sec., light—4 sec. beginning 1 sec. after the tone, and darkness—25 sec., consumed just 30 sec. and was repeated twice a minute for 25 min., making a total of 50 trials. On Trial 51 the tone came on as usual, but instead of the light following 1 sec. later, the camera and infrared lamps followed at this interval, and three pictures a second were taken for approximately 10 sec.

The camera and infrared lights were connected to the same timer. Since the two are not integrated exactly, 30 pictures were not taken each time. However, all *S*s had 28 pictures, some had 29, and fewer had 30; for this reason the data for 28 pictures at each test are presented in Fig. 1.

Before beginning the conditioning trials a preliminary test was made using the white light, but no tone, to determine the nature of *S*'s contraction to the UCS. This was followed 25 sec. later by a second preliminary test with the tone, but without the white light, to determine whether there was any response to the tone. Immediately following these two preliminary trials, which were given only once at the beginning of the experiment, *S* was presented the first set of training trials. All time sequences were preset, and the entire procedure ran off automatically once it was started.

Results

The results for males and females are presented separately in Fig. 1 for comparative purposes. These results were subjected to an analysis of variance to determine whether or not there was a significant difference between males and females, or between trials, since a contraction of the pupil on a test should result in a significant difference between the measurements made on that test. The test for differences between trials was made for males and females separately and combined. The results of this analysis are presented in Table 1.

An examination of the curve for Preliminary Test 1 in Fig. 1 reveals that both males and females show contractions to the UCS of approximately 3.25 mm., with the course of contraction being very similar for both groups. On the earlier trials the females tend to have larger pupil diameters than the males. They also show a small contraction on all trials except Trials 5, 6, and 7. There is no evidence of the development of a conditioned pupillary response which would fit the description given by Hudgins (5).

The analysis indicated that the differences observed in Fig. 1 between males and females do not even approach significance. Some of the dif-

TABLE 1

Analyses of Variance between Trials of Each Conditioning Test for Males and Females Separately and Combined

Test	Males		Females		Both	
	F	*df**	*F*	*df**	*F*	*df**
Pre. 1	64.642	221	79.356	236	142.815	457
Pre. 2	2.210	233	1.037	242	9.520	476
1	1.054	237	1.131	223	3.086	460
2	<1	219	1.201	235	1.376	435
3	5.660	217	<1	207	2.479	424
4	1.131	222	3.504	211	3.094	434
5	4.363	222	2.285	233	<1	456
6	6.734	238	4.304	239	9.758	475
7	1.884	217	1.169	201	1.915	418
8	3.599	147	4.524	228	<1	375

* Degrees of freedom for error variance; *df* for numerator is 27 in all cases. The .05 and .01 values of *F* for 27 and 200 *df* are 1.55 and 1.84, respectively; for 27 and 400 *df*, 1.52 and 1.79.

ferences between, and the variability of, the curves for males and females may be explained by the varying number of *Ss'* scores occurring on different trials. This variation in number of scores resulted from difficulties with the camera, with the developing process, and with *Ss* who blinked, moved, etc. Since 28 pictures were to have been taken of each of 20 *Ss*, the maximum number of pictures for a single test is 560. Two of the tests, the second preliminary test and Test 6, had a total of 550 or more pictures; the first preliminary test and Tests 1, 2, and 5 had totals between 525 and 550; Tests 3, 4, and 7, between 500 and 525; and Test 8 had 450 pictures. The varying number of scores would fail to explain the largest reversal of the males and females, which occurs in Test 6, since only ten pictures which were equally distributed between the two groups were lost in this test. A review of our procedure did not indicate any other explanation of the differences between the two sets of data.

The small contraction clearly demonstrated for the females in 6 out of 9 trials and for the males in 2 out of 9 trials led to an analysis of the difference between trials to determine whether or not this was significant. Since no sex differences were found, the analysis was made for the males and females separately and combined. Only on the first preliminary test were these three tests in consistent agreement. The largest contractions for the female *Ss* occur in Tests 2, 4, and 8. Of these three tests, Test 2 shows no significant difference, while Tests 4 and 8 yield highly significant differences, but they also show much more variation throughout the entire curve. The first and second tests which show contractions for the males show no significant differences between pictures. Therefore, it appears that there is no clearly significant pupillary contraction occurring in any of the tests. This early and small contraction, which occurs most often with female *Ss*, may be similar to the psychosensory contraction discussed in an earlier paper (13). Since the contraction is followed by a dilation within a fraction of a second, it may be a reaction to the onset of the infrared lights and the camera. Further, because the contractions were followed by a dilation in almost all cases, they are not due to any visible light component of the infrared lights since there would be no redilation in this case.

Discussion

A comparison between these contractions and the pupillary conditioned contractions described by Hudgins may be made. Hudgins (5) reported a latency of 2 sec. and a contraction duration of 11 sec. In all the contractions shown in Fig. 1, the latency is less than .33 sec., and the contraction duration varies between .66 sec. and 1.33 sec. Thus, all of the contraction is completed before Hudgins' contraction could have begun. The latencies and durations of contractions found in this study agree quite well with those found in the earlier study (13).

Since the results of this study do not support those of Cason and Hudgins, the possible reasons for this failure should be examined, particularly with respect to Hudgins' study. The methodology of the present study differs from that used by Hudgins (5) as follows: (*a*) Fewer training trials were run at each session and more test trials were made. (*b*) A bite board was used to keep *S* in focus with the camera. (*c*) The CS preceded the onset of the UCS by 1 sec. instead of being simultaneous. (*d*) The test trials were made in visual darkness instead of dim visible illumination. (*e*) A fixation point was used throughout the training trials. (*f*) A camera was used instead of a telescope.

The use of fewer training trials at each session should have lessened any possibility of fatigue developing that might mask the CR before the test trials were made. A further check on this factor is provided by a preliminary training series run on one *S*. In this series *S* was given 200 training trials in groups of 25 trials with test trials being made after Trials 1, 5, 10, 25, 50, 75, 100, 125, 150, 175, and 200. The results of this series were the same as those obtained in the present study. Since Hudgins obtained conditioning after longer sessions, the variations employed here should have heightened the CR if fatigue may be considered as a depressor of the response.

The bite board used in conjunction with a headrest in the present study restricted head movements, but no restrictions were placed on the remainder of the body, so that *S*s were able to adjust the stool and their positions in order to be relatively comfortable. Although each *S* was asked about the discomforts of the situation, no one reported that it was uncomfortable, but some complained that they had difficulty trying to stay awake. In this connection it is interesting to compare the situation of head restriction employed here with the situation used by Hudgins. A duplicate of his apparatus was built to check several other differences that will be described later. Hudgins makes no mention of any head support or bite board, but in his apparatus any movement in any direction makes it almost impossible to retain fixation on the cross hairs. If there is any strain that may be related to fixation, it is much greater with his apparatus, if no mechanical head support was used in his experiment.

The presentation of the CS 1 sec. before the UCS in the present study represents somewhat of a compromise. Since the other investigators (**4, 8, 9, 10, 11, 14**) who repeated Hudgins' study used a simultaneous presentation with no success, since Baker (1) obtained conditioning in two trials when the CS began 9 sec. before the UCS, and since a slightly prior presentation of the CS generally yields better conditioning, it was felt that an interval of 1 sec. between the CS and UCS might be more favorable to learning.

The use of darkness instead of dim illumination and some pupil contraction during the test trials should have yielded a larger magnitude of conditioned pupillary contraction if such were present.

The variations presented above should have presented more favorable conditions for conditioning taking place and should not have caused the present failure to show conditioning. The use of a constant fixation point and the use of a camera instead of a telescope may be responsible for the differences in results. The task of aligning hair lines with the edges of the fluctuating pupil while using indirect vision have already been discussed. This task is subject to so much individual judgment and variability that trends in the data might conceivably be introduced by *E*'s expectations, however careful he might be regarding this source of error.

In the Hudgins experiment two variables are related to pupillary contraction, the light and the change in accommodation. In the present experiment the change in accommodation was removed by having *S*s maintain constant accommodation throughout the experimental session. This difference in procedure may have affected the results in two ways: it may have resulted in better conditioning, or it may have suppressed conditioning. It is impossible to say what effect the constant accommodation has had in the present experiment. Hudgins felt that his results were not due to variations in accommodation in the test situation, since he found with his apparatus that a 4-diopter change in accommodation resulted in a .31-mm. change in pupil size. When a duplicate of his apparatus and his test conditions was used, no change in *S*s' accommodation was found on retinoscopic examination. In all probability there was no change in accommodation in the test situation which would explain his results, but this does not rule out the possibility of the occurrence of accommodation during conditioning trials affecting the CR. The .31-mm. change in pupil diameter with a 4-diopter change

in accommodation does not agree with results found by more recent investigators using photographic techniques. Marg and Morgan (7) found a .30-mm. change in diameter for each diopter change in accommodation. This yields a total value of 1.20 mm. instead of .31 mm. for the distances used by Hudgins. This difference may also be due to differences in measurement techniques.

Summary

A photographic study of pupillary contraction conditioning was made using ten male and ten female *Ss*. The experimental conditions used were similar to those employed by Cason and Hudgins. Three pictures per second were taken for approximately 10 sec. using infrared illumination. No evidence of pupillary conditioning was found. The results of the portions of Cason's and Hudgins' experiments dealing with the conditioning of the pupillary contraction to a sound stimulus were not substantiated.

References

1. Baker, L. E. The pupillary response conditioned to subliminal stimuli. *Psychol. Monogr.*, 1938, **50**, No. 3 (Whole No. 223).
2. Cason, H. The conditioned pupillary reaction. *J. exp. Psychol.*, 1922, **5**, 108–146.
3. Hilgard, E. R., Miller, J., & Ohlson, J. A. Three attempts to secure pupillary conditioning to auditory stimuli near the absolute threshold. *J. exp. Psychol.*, 1941, **29**, 89–103.
4. Hilgard, E. R., Dutton, C. E., & Helmick, J. S. Attempted pupillary conditioning at four stimulus intervals. *J. exp. Psychol.*, 1949, **39**, 683–689.
5. Hudgins, C. V. Conditioning and the voluntary control of the pupillary light reflex. *J. gen. Psychol.*, 1933, **8**, 3–51.
6. Lowenstein, O., & Friedman, E. D. Pupillographic studies. I. Present state of pupillography; its method and diagnostic significance. *Arch. Ophthal.*, 1942, **29**, 969–993.
7. Marg, E., & Morgan, M. W., Jr. The pupillary near reflex; the relation of pupillary diameter to accommodation and the various components of convergence. *Amer. J. Optom.*, 1949, **26**, 183–198.
8. Steckle, L. C. Two additional attempts to condition the pupillary reflex. *J. gen. Psychol.*, 1936, **15**, 369–377.
9. Steckle, L. C., & Renshaw, S. An investigation of the conditioned iridic reflex. *J. gen. Psychol.*, 1934, **11**, 3–23.
10. Stern, F. An investigation of pupillary conditioning. Unpublished doctor's dissertation, Univ. of Washington, 1948.
11. Wedell, C. H., Taylor, F. V., & Skolnick, A. An attempt to condition the pupillary response. *J. exp. Psychol.*, 1940, **27**, 517–531.
12. Young, F. A., & Biersdorf, W. R. An apparatus for taking rapid photographs of the pupil in visual darkness. *Amer. J. Psychol.*, 1952, **65**, 617–618.
13. Young, F. A., & Biersdorf, W. R. Pupillary contraction and dilation in light and darkness. *J. comp. physiol. Psychol.*, 1954, **47**, 264–268.
14. Zabarenko, R. N. The conditioning of the pupillary light reflex. Unpublished master's thesis, Univ. of Pittsburgh, 1939.

(Received September 23, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

THE RELIABILITY OF MEASUREMENTS OF HUMAN DARK ADAPTATION [1]

F. A. MOTE
University of Wisconsin

G. E. BRIGGS
Northwestern University

AND K. M. MICHELS
Purdue University

This experiment was undertaken for several reasons. One was to obtain information upon the day-to-day variability of measurements of dark adaptation in the human eye over a period of several months. Most experiments upon the consistency of dark-adaptation data have involved only two or three repetitions of measurements on the same *Ss* (2); we wished to procure more measurements, and over a longer course of time. Another reason was to gather data upon the variation of two immediately successive sets of dark-adaptation measurements on the same *S* for the same conditions of pre-exposure. A third reason was to obtain data upon any differences between experimental sessions conducted in the forenoon and those conducted in the afternoon. Finally, we wished to know, for the same *S* and for the same conditions of pre-exposure, whether or not having different *Es* would introduce differences in the dark-adaptation measurements.

Apparatus and Procedure

The apparatus used both for pre-exposure and threshold measurements was a Hecht-Shlaer adaptometer. The pre-exposure light was white, and its intensity was controlled by a Wratten neutral tint fixed density filter. When threshold measurements were taken, a violet filter was used with neutral tint fixed filters so that *S* could report on the presence of color at threshold. The threshold testing stimuli were presented by a pendulum-type shutter attached to the adaptometer, and their duration was .2 sec. The pre-exposure patch subtended 35° visual angle, the threshold test patch, 5°. The right eye was used, and both pre-exposure and test stimuli were imaged 7° from the fovea on the temporal portion of the retina. No artificial pupil was used.

Each session was conducted as follows: The *S* entered the light-tight experimental cubicle, and the pre-exposure was presented at once; there was no standardization of pre-experimental conditions. At the expiration of the pre-exposure, the initial threshold was determined as quickly as possible, usually within 7–15 sec. Threshold determinations were taken for 40 min., every half-minute for the first 5 min., then every minute. The same intensity and duration of pre-exposure were given again, and the testing procedure repeated. Two dark-adaptation curves, each of 40 min. duration, were thus obtained. A complete session required about 1.5 hr.

To obtain a threshold measurement, *S* was first presented an intensity slightly higher than his previous threshold, a faintly visible stimulus. The intensity was then reduced to below threshold, and the threshold value approached by increasing the intensity by steps of .01 to .03 logarithmic units until the stimulus was reported as visible for two successive presentations of the same intensity.

The pre-exposure intensity was 1,244 mL., and its duration was 2 min. This degree of light adaptation was not so high that the final level of dark adaptation could not easily be attained in 40 min. Consequently, the time required to obtain the two curves per session did not fatigue the *S*. It was sufficiently high, however, to elevate the threshold enough to provide both rod and cone components to the dark-adaptation curves.

For each *S* there were 12 experimental sessions, irregularly spaced over a period of 11 months. Since two sets of measurements were obtained at each session, a total of 24 curves was obtained for each *S*. Half the sessions were in

[1] This research was supported by funds from the Research Committee of the Graduate School of the University of Wisconsin and by a contract between the Office of Naval Research and the University of Wisconsin.

the forenoon; the other half in the afternoon or early evening. For each *S* half the sessions were conducted by each of the two *E*s, and the schedule was so arranged that each *E* conducted half his sessions in the forenoon and half in the afternoon. Thus, the data obtained provided information for a given *S* from two sets of measurements taken in immediate succession, and from day to day over the course of several months, as well as for different times of day and with different *E*s. One of the *S*s (FAM) was experienced; his dark adaptation has been measured several hundred times during the past few years. Another *S* (GEB) was less experienced; he had served in an experiment in which his dark adaptation had been measured about fifty times. The third *S* (KMM) was inexperienced.

Results

In Fig. 1 are shown the means and the range of ± 1 *SD* for KMM for the two curves obtained at each session. The raw data were recorded in log micromicrolamberts. The 12 sets of values for the first curves of each session were added and divided by 12, and the same was done for the 12 sets for the second curves of the sessions. The data thus obtained were geometric means. The standard deviations are based upon these means.

The curves shown in Fig. 1 are representative of the two sets of curves per session; the data for the other *S*s show essentially the same features. The figure exhibits the classical pic-

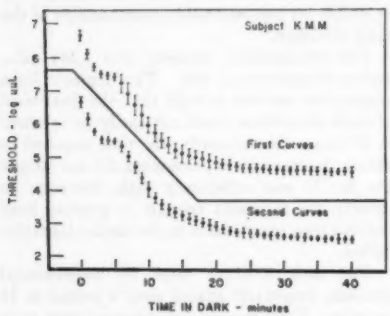


Fig. 1. Mean dark-adaptation threshold values and the limits of ± 1 *SD* for KMM

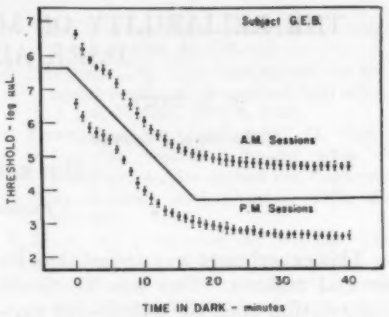


Fig. 2. Mean dark-adaptation threshold values and the limits of ± 1 *SD* for GEB

ture of the course of dark adaptation following a moderate degree of light-adapting pre-exposure. The total extent of dark adaptation is about 4 logarithmic units. Each curve shows a rapid descent from an initial threshold value of around 6.6 log micromicrolamberts, and the descent continues for 3 or 4 min. For the next 2 or 3 min. the threshold values drop only slightly, indicating that the end of the period of cone adaptation is approaching. The extent of this cone component of the curve is approximately 1.5 logarithmic units. At around 5–7 min. a second rapid descent begins. This "break" in the curve represents the beginning of rod adaptation. Also, within a minute or two of this time, the character of the threshold stimuli changes. Up to this time they have been violet; now they are colorless and, of course, continue to be for the remainder of the period of dark adaptation. The extent of rod adaptation is about 2.5 logarithmic units, the final threshold value for each *S* being in the neighborhood of 2.6 log micromicrolamberts.

In Fig. 2 are shown the means and the range of ± 1 *SD* for GEB for the morning and afternoon sessions. In Fig. 3 the same data are presented for FAM divided according to *E*s. These

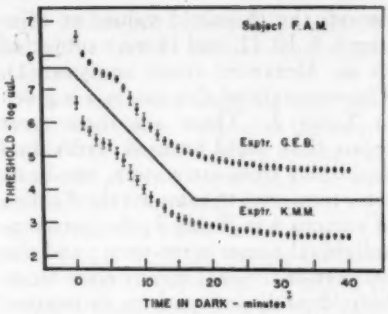


FIG. 3. Mean dark-adaptation threshold values and the limits of ± 1 *SD* for FAM

are very similar to Fig. 1. Among the figures there is an obvious difference in the magnitude of the *SD*'s; the values are largest in Fig. 1 and smallest in Fig. 3. Some index of the difference can be obtained from the mean *SD*'s. For Fig. 1 the mean *SD* for the data in the upper panel is .16 log micromicrolambert (σ_{σ} = .018), and in the lower panel .12 log micromicrolambert (σ_{σ} = .013). For Fig. 2 they are .12 (σ_{σ} = .013) and .11 (σ_{σ} = .012), and for Fig. 3, .08 (σ_{σ}-.009) and .09 (σ_{σ}-.010).

Several statistical analyses were made in order to provide answers to the following questions: (*a*) Will significant variability have been introduced into the data because of differences in the time of day at which the sessions were conducted; because one curve was determined immediately following another in each session; and because each *S* was tested by two *E*s, thereby introducing an experimenter-subject variability? (*b*) Will any such variability as may be found associated with these factors be a feature of the whole course of dark adaptation, or only of certain stages? (*c*) Is there a practice effect extending over the duration of the experiment?

The sections of the dark-adaptation curve selected for these analyses were: the initial threshold values, the thresholds at the rod-cone "break," the course of the threshold during the early part of rod adaptation, the last six threshold values, and the values at the fortieth minute.

Table 1 presents a summary of an analysis of variance for the initial threshold values. The *E–S* factor was found to be significant at the 5% level of confidence. A similar analysis showed the same level of significance for this factor for the values at the rod-cone "break."

In the analysis of Table 1, the residual mean square was used to test each of the main factors (first or second curve in a session, forenoon versus afternoon sessions, and *E–S* combination) and their interactions. Only the *E–S* factor was found to be significant. The second-order interaction was also significant when tested against the residual-error mean square; therefore a pooled error term was computed in order to further test the significance of this factor. Since it was found to be still significant at the 5% level of confidence, we may conclude that it was a source of variation beyond that attributable either to random variability or to interaction variability, or to both.

A further analysis of the initial threshold values, in which the effects of *S*s and *E*s were separately investi-

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE FOR INITIAL THRESHOLD VALUES

Source	*df*	Mean Square	*F*	F_p
Experimenter-Subject (*E-S*)	5	.0924	3.51*	3.31*
Curve 1–Curve 2 (1–2)	1	.0019		
AM-PM	1	.0861	3.27	3.09
E-S ×1–2	5	.0106		
E-S ×AM-PM	5	.0041		
AM-PM ×1–2	1	.0416	1.58	
E-S ×AM-PM ×1–2	5	.0663	2.52*	
Within	48	.0263		
Total	71			

* Significant at the .05 level of confidence, or better.

gated, revealed that the *E–S* factor was significant only when GEB and KMM were *E*s; there was no significance attributable to *S*s. No significance was found when FAM was either *S* or *E*. This result indicates that experience in dark-adaptation experimentation is probably the most influential feature contributing to the significant *E–S* factor. The same sort of analysis applied to the values at the rod-cone "break" showed *S* to be the significant effect; *E*s, not significant. This finding is merely the expression of the fact that individuals differ in the threshold level at which the adaptation of the cones is superseded by that of the rods.

The analyses concerned with the factors of first or second curve in a session and forenoon versus afternoon sessions showed these to be of no significance.

None of the factors or interactions was found to be significant for the threshold value at the fortieth minute.

The portion of the dark-adaptation curve that covers the largest extent of threshold values, and the portion showing the greatest variability, is the rapid, beginning phase of rod adaptation. In order to investigate the course of dark adaptation during this period, the threshold values at Minutes 6, 8, 10, 12, and 14 were subjected to an Alexander trend analysis (1). The summary of this analysis is given in Table 2. There are three error terms that could be used: individual deviations from estimation, the basic error term used to compute the *F* ratios of column 4 in Table 2; the between-individual slopes error term; and the between-individual means error term. Individual deviations from estimation is an estimate of pure error and may be assumed to be a good estimate of such random error as entered into the results of the experiment. When between-group means and between-group slopes are tested against this basic error term, highly significant differences are found. However, when these two additional error terms are used, the *F* ratios—obtained by dividing the between-groups means by the between-individual means, and the between-groups slopes by the between-individual slopes (column 5, Table 2)—are no longer significant. Thus, the separate significance or nonsignificance of the three main factors and their interactions cannot be validly determined for the rapid, beginning phase of rod adaptation. The day-to-day variability accounts for any signifi-

TABLE 2

ALEXANDER TREND TEST FOR THRESHOLD VALUES DURING THE EARLY PHASE OF ROD DARK ADAPTATION

Source	*df*	Mean Square	F_{IDFE}	*F*
Between-group slopes	23	.0186	93.00**	
Between-group means	23	.0535	267.50**	1.30
Between-individual slopes	48	.0219	109.50**	
Between-individual means	48	.0410	205.00**	
Over-all slope	1	57.4232	287116.00**	
Over-all deviation from linearity	3	.2257	1128.50**	
Group deviations from estimation	69	.0042	21.00**	
Individual deviations from estimation	144	.0002		
Total	359			

** Significant at .001 level of confidence, or better.

cance which might be attributable to the three main factors and their interactions.

A second trend analysis was performed in order to determine the existence of any practice effect during the 11 months of experimentation. It was assumed that the results of any learning during the experiment would be most readily detected during the final phase of dark adaptation. Figures 1, 2, and 3 show that it is at this time that the threshold approaches a stable, limiting, final value, and it is also during this period that the variability is the smallest. For the analysis, we therefore selected a measure of variability for the six threshold values from the thirty-fifth to the fortieth minute, inclusive. Six sessions per *S*, approximately equally spaced in time throughout the experiment, were analyzed. In neither the data for the individual *S*s nor in the data for the *S*s taken as a group was there any evidence of a practice effect.

In summary, the results of the statistical analyses reveal that, of the three main factors investigated (*E*–*S* combination, first or second curve in a session, and forenoon versus afternoon sessions), only the *E*–*S* factor was found to be significant. This was significant for the differences among the initial threshold values and the values at the rod-cone "break." No factor was found to be a significant source of variation at the final threshold value, and the influence of any factors during the early, rapid phase of rod adaptation was indeterminate because of day-to-day variability among the individual dark-adaptation curves. No practice effect was found.

Discussion

The variability of threshold measurements changes during the course of dark adaptation. This is most clearly seen in Fig. 1, but it is also apparent in the other two figures. For the cone component, the spread is largest for the beginning two or three threshold determinations. Since any aftereffects of the pre-exposure would be most prominent during this period, and also because the threshold is falling very rapidly, it is to be expected that the values would show greater variability during this early part of cone adaptation. With the approach of the end of this component of the curve, the variability becomes smaller and remains so until rod adaptation begins. Then it again increases. There are two reasons for this. One is the same reason as that offered to account for the greater variability at the early part of the cone component: during this phase the threshold of the rods is falling rapidly; consequently a difference of a few seconds in the time of measuring the threshold at a given minute, from session to session, makes a considerable difference in the values. The other reason is the variation in the time at which rod adaptation begins. For example, at one session the rod component may start at 5 min.; at the next, at 7 min. Since the drop in rod threshold may easily be half a logarithmic unit in 2 min. during this early phase of rod adaptation, a time difference of this amount for the beginning of the rod component may make a marked difference in the threshold values at a given instant from one session to another. This relatively large variability continues, but tends to become smaller, until about the twentieth minute. During this time, the threshold has descended approximately two logarithmic units. The descent from then until the fortieth minute is only about one-half a logarithmic unit, and the variability remains at a low, fairly stable value.

In many studies *E* finds it advantageous to obtain data for different values of the independent variable, on the same *S*, at a single session. The question then arises: Does the fact that a curve of dark adaptation has just been obtained have any influence upon the measurements for

a curve to be obtained immediately afterward? With data available on the variation of two sets of measurements taken at a single session for a fixed value of the independent variable, we have a criterion for assessing the effects of changing the independent variable. The results of this experiment show that no detectable effect is found in a set of measurements that can be accounted for by the fact that another set has just been obtained in the same session. The same is true for sets of measurements taken at different times of the day. In studies of dark adaptation *E* may therefore feel secure that any obtained differences are not inflated because the measurements have been taken in immediate succession or at different periods during the day.

Summary

Measurements of dark adaptation were made with a Hecht-Shlaer adaptometer following preexposures to 1,244 mL. intensity for 2-min. duration. The experiment was carried out over a period of 11 months, the authors serving as *E*s and *S*s. For each *S* a total of 24 dark-adaptation curves was obtained, two curves of 40-min. duration per session for 12 sessions. Statistical analyses were carried out to determine the influence of several factors for the initial threshold value, the value at the rod-cone "break," and the final threshold values.

The main results were as follows:

1. For each *S* the total extent of change in threshold values for each curve was approximately 4 logarithmic units. All curves clearly showed both cone and rod dark-adaptation components.

2. No significance at any stage in dark adaptation was found to be associated with the factors of forenoon versus afternoon, or first curve versus second curve in a session.

3. The factor of experimenter-subject combination was found to be significant at the 5% level of confidence for the initial threshold values and for the values at the rod-cone "break." Further analysis revealed that for the initial threshold values, this is not to be attributed to the *S*s, but to two of the *E*s, and is attributable to variation in experience in taking measurements. The significance found for the threshold values at the rod-cone "break" was associated with *S*s, not *E*s, and is an expression of the fact that individuals differ in the threshold level at which the adaptation of the cones is replaced by that of the rods.

4. The influence of any factors during the early phase of rod adaptation could not be determined because of day-to-day variability among the individual dark-adaptation curves.

5. No practice effect was found.

References

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. Sheard, C. Dark adaptation: some physical, physiological, clinical, and aeromedical considerations. *J. opt. Soc. Amer.*, 1944, **34**, 464.

(Received November 6, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

EFFECT OF NUMBER OF ALTERNATIVES AND SET ON THE VISUAL DISCRIMINATION OF NUMERALS [1]

GILBERT K. KRULEE, JEROME E. PODELL,[2] AND PAUL G. RONCO [3]

Tufts College

The experiments to be described in this report have as their main emphasis the determination of relationships between amount of information involved in a particular visual discrimination and the ease of making such a discrimination. Two hypotheses were advanced: (*a*) that the magnitude of the threshold for discrimination of numerals would increase as the number of alternative possibilities (amount of uncertainty) was increased, and (*b*) that in a situation in which expected uncertainty exceeded actual uncertainty, the visual threshold would be a function of the number of alternatives expected by *S* and not of the number actually used by *E*. These hypotheses have considerable similarity to those advanced in certain recent studies on reaction time as a function of amount of information processed (**2, 3, 4, 5**).

METHOD

Apparatus.—An experimental apparatus was developed so that numerals could be displayed and the distance threshold for recognition of these numerals be determined. The apparatus consisted of a box 35½ in. long, 6 in. by 7 in. in cross-section, containing a fixed eyepiece at one end and enclosed by a frosted glass screen at the other. The interior of the box was illuminated only by light from two shielded 25-w. lamps fastened to the box. Behind the screen was located a movable carriage by means of which numbers could be displayed. The carriage was driven by a constant-speed electric motor (1/150 hp, 1800 rpm) through a gearing arrangement. For each presentation of a number, the display was initially positioned far enough away so that neither the carriage nor the number displayed was discernible. The *S* pressed a switch to activate the carriage motor, thus driving the carriage in toward the glass screen. The instructions to *S* were to release the switch as soon as he could recognize the number displayed so as to stop the travel of the carriage. The threshold value was defined as the amount of travel required from the initial starting point before recognition was possible. This threshold was measured by a counter which recorded in tenths of a revolution of the drive shaft, 21.2 revolutions being equivalent to 1 in. of movement along *S*'s line of sight. Subsequently, these threshold values were converted into distance in inches from the eye of *S* and are so recorded in the appropriate tables of results.

The display itself consisted of three wheels mounted at right angles to the line of sight, by means of which any 1-, 2-, or 3-digit number could be displayed. Each wheel had 12 faces, 2 of which were blank, while the remainder contained the set of all one-digit numbers in order, one number to a face. In order to hold the wheels in place for the display of a given number, a spring-loaded window was provided.

The numerals employed are those of Berger (1), which are equated for relative ease of discrimination. The height of the numerals was 1 in. They were painted in white on a black background, with all other areas of the associated display in black, including the blank positions.

Procedure.—The experimental sessions were always begun with some general instructions by *E* as to the nature of the experiment and the operation of the apparatus. The *S* controlled the motion of the display by means of a pressure-operated switch held in his hand.

Before each presentation, *E* moved the display carriage back to the predetermined zero position. After *E* had set the wheels so as to display a given number and had recorded the initial

[1] The experiments upon which this report is based are supported by a contract between the Office of Naval Research and Tufts College, and are part of a long-term program of research on the capabilities of humans as processors of information. The authors wish to express their appreciation to Walter Silvester, Richard H. Lent, and Paul T. Perney, all of Tufts College, for their assistance in the prosecution of this work.

[2] Presently located at Clark University.

[3] Presently on active duty with the U. S. Army.

counter reading, *S* was free to start the motor at his own convenience. His instructions were to position his head by means of the eyepiece and to continue without interruption the movement of the display; as soon as he could recognize the number displayed, he was to simultaneously release the switch (thus stopping the motor) and give his response verbally. Throughout, instructions emphasized accuracy, although *S* was also informed to give his response as soon as he was reasonably certain of his decision. After each presentation, *E* recorded the response and the counter reading and made preparations for the presentation of the next element in the sequence. Presentations were continued without knowledge of results as to threshold values or accuracy being given. Threshold values for inaccurate responses were always omitted from calculations and were replaced by the value obtained from an additional presentation.

Experiment I

The first experiment conducted with the previously described apparatus involved the determination of this distance threshold under variations in the number of alternatives from which a random selection of presentations was made.

Method.—The experiment was carried out in two parts, each of which employed 32 *Ss*. The *Ss* in both this and the following experiment were male naval enlisted personnel obtained through the cooperation of the U. S. Naval Receiving Station, Boston, Massachusetts. They were selected without reference to any particular qualifications.

In the first part of the experiment, *Ss* read numbers under conditions of 2, 4, 8, and 16 alternative categories. Within each condition, *Ss* were given a sequence of four randomly chosen selections with replacement, with full knowledge to *S* of the set of possible alternatives. The analysis of the data was based upon the average of these four thresholds. Rotation of conditions was introduced in order to control for possible practice effects. This rotation was carried out so as to produce two different sets of latin squares, one set of the form ABCD, BCDA, CDAB, DABC, and the other of the form DCBA, CBAD, BADC, ADCB. The only change introduced into the second half of the experiment was that the conditions of information used were 16, 32, 100, and 1000 categories.

In both parts of the experiment, consideration was given to ensure that any differences in the relative ease of discrimination of particular numbers would be controlled. This was accomplished by a design that was balanced as to the frequency of occurrence of any specific number. For each *S*, a different random sequence of presentations was selected. In addition, for the first half of the experiment, the scheme of balancing summarized in Table 1 was followed. The use of this method was chosen with two objectives in mind: (*a*) considering all *Ss* within a given informational condition, to ensure that each number had an equal probability of occurrence; and (*b*) for any given *S*, to ensure that the sequence of presentations was not unduly weighted by the presence of any specific number. For the second half of the experiment, similar objectives were met by

TABLE 1

Categories Selected for Presentation

Number of Categories	Subgroups*															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	15 16	17 18	21 22	23 24	25 26	27 28	31 32	33 34	35 36	37 38	41 42	43 44	45 46	47 48	11 12	13 14
4	21–24		25–28		31–34		35–38		41–44		45–48		11–14		15–18	
8	31–38				41–48				11–18				21–28			
16	11–18 21–28								31–38 41–48							

* There are 16 different subgroups, with two *Ss* in each subgroup.

TABLE 2

DISTANCE THRESHOLD VS. INFORMATION PROCESSED

Number of Categories	Mean	*SD*	Paired Differences		*t*
2	40.09	.386			
4	40.00	.208	A-B	.09	2.52*
8	39.92	.197	B-C	.08	2.86*
16	39.93	.163	C-D	−.01	.53
16	39.91	.263			
32	39.90	.342	E-F	.01	.90
100	39.88	.204	F-G	.02	.98
1,000	39.90	.169	G-H	−.02	1.41

* Significant at .01 level of confidence, based on 31 *df*.

providing the following arrangement: For the condition of 16 categories, half of the *S*s received sequences chosen from the set 11–18, 21–28, with the other half receiving choices from the set 31–38, 41–48. For 32 categories, the set of possibilities used throughout was 11–18, 21–28, 31–38, and 41–48; for 100 categories, 00–99; and for 1,000 categories, 000–999.

Results.—The results are summarized in Table 2. In order to study the effect on threshold values of increases in information, scores were paired for a given *S* under the four experimental conditions, and the *t* test was applied to these differences. From the results, it is apparent that there is a significant increase in threshold as the amount of information is increased from 2 to 4 and from 4 to 8 categories. No further increase in threshold can be detected for amounts of information in excess of 8 categories.

In addition, an analysis of variance was carried out on each half of the experiment, primarily to investigate whether or not any practice effect could be detected. The results of this analysis are summarized in Table 3. The procedure for this analysis of variance was based upon a method suggested by Mann (**6**, p. 80) and by Tippett (**7**, p. 231) for the analysis of a series of latin squares. In this case the experimental data could be summarized in terms of eight latin squares, four of which had the form based upon the order of presentation ABCD, BCDA, etc., while the remaining four were based upon the order of presentation DCBA, CBAD, etc. The effect of the variation due to the difference in order of presentation is listed in Table 3 under the heading "Variations in presentation." As noted in Table 3, the null hypothesis concerning practice cannot be rejected at either the 1% or the 5% level of confidence in either case. Under the conditions of this experiment, *S*s did not seem to improve their performance as a result of practice.

The results for the variable of amount of information corroborate the findings reached by the use of the *t* test. The variation due to information is significant at the 1% level of confidence for the first half of the experiment, but is not significant even at the 5% level of confidence for the second half.

EXPERIMENT II

The design of the second experiment was chosen so as to investigate

TABLE 3

ANALYSIS OF VARIANCE OF THRESHOLD MEASURES

Exp.	Source of Variance	*df*	Mean Square	*F*
I	Information	3	8640	13.3*
	Subjects	31	4594	7.09*
	Practice	3	715	1.10
	Variations in presentation	1	4864	7.51*
	Error	89	648	
	Total	127	1834	
II	Information	3	330	1.38
	Subjects	31	3720	15.6*
	Practice	3	128	.54
	Variations in presentation	1	2496	10.4*
	Error	89	239	
	Total	127	1106	

* Significant at .01 level of confidence.

TABLE 4

DISTANCE THRESHOLD UNDER RESTRICTED KNOWLEDGE OF THE APPROPRIATE SET

Conditions	I		II		III	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
A. Total set	39.92	.208	39.89	.079	39.94	.120
B. Restricted set without knowledge	39.97	.175	39.98	.147	39.99	.190
C. Restricted set with knowledge	40.15	.182	40.18	.270	40.24	.221

Analysis of Paired Differences

Conditions		*t*	Conditions		*t*	Conditions		*t*
I_B-I_A	.05	1.05	I_C-I_B	.18	4.08**	I_A-II_A	.03	.49
II_B-II_A	.09	2.59*	II_C-II_B	.20	3.77**	II_A-III_A	−.05	1.48
III_B-III_A	.05	1.04	III_C-III_B	.25	3.46**			
						I_C-II_C	−.03	.50
						II_C-III_C	−.06	.95

* $p = .05$, $df = 11$.
** $p = .01$, $df = 11$.

the consequences of perceived uncertainty being greater than objective uncertainty. As part of the experimental design, an effort was made to establish a perceptual set on the part of *S* to expect more categories than were actually possible. It was assumed that, while *S*s were operating in terms of this expectation, their threshold values would be higher than would be attained if they had more adequate knowledge as to the set of categories from which alternatives were being chosen.

Method.—The experiment consisted of three variations, each of which was administered to the 12 *S*s taking part in the experiment. In the first variation, *S* was shown a sequence of 12 numbers chosen in the following fashion: The first four numbers were chosen randomly with replacement from the set of the one-digit numbers. In order to select the rest of the numbers to be presented, a subset of four numbers from the set of one-digit numbers was chosen, and the possibilities for the remaining elements in the sequence were restricted to this subset. Then eight numbers were chosen on a random basis from this restricted set of four possible alternatives.

As to the actual presentation of the 12 elements in the sequence, the first four numbers in the sequence were displayed to *S* without any special instructions. It was assumed that the expectation would be established on the part of *S* that the set of possibilities from which choices were made was that of all one-digit numbers. When the second series of four presentations was given without any interruption or further instructions, it would seem reasonable to assume that the previously established expectation would be carried over and would lead *S* to behave as if more information was being transmitted than was actually the case. He would act as if there were ten possibilities, whereas, in fact, there were only four. After the presentation of this second series of four numbers, *S* would be told that for all subsequent presentations, only four values were possible, and he would be given these specific values. Thus, the final four presentations would be given with complete knowledge to *S* of the set of possibilities from which they were being chosen.

This same design was repeated with the 100 two-digit numbers and the 1,000 three-digit numbers. In each variation, *S*s initially received no instructions as to the set of possibilities for the first eight presentations. The first four were always chosen randomly with replacement from the total set of either 1-, 2-, or 3-digit numbers in order to establish the expectation that this was the set from which choices were being made. The second series of four numbers

was chosen from a restricted set of four possibilities selected randomly from the total sets, and this series was presented without additional instructions. The final series of four presentations was then given with full knowledge to *S* of the set of four values to which the possibilities were restricted.

The order of presentations of the three variations was balanced by a rotation of variations of the form ABC, BCA, and CAB. Since no significant practice effects had been obtained in the first experiment, no attempt was made to rotate conditions within a given variation. Different random sequences were selected for each *S* and the restricted subsets were rechosen for each *S* so as to balance out any effects due to unequal ease of discrimination of particular numbers.

Results.—Analysis again was based on the *t* test applied to paired scores for each *S*. The results are summarized in Table 4.

On the basis of the results, a number of comparisons can be made. One such comparison is between the average of the second set of four presentations as compared to the third set of four within each variation (i.e., the average threshold for the restricted subset without knowledge vs. the value for the same subset with knowledge of its composition). For all variations, there are sharp drops in thresholds, all significant at the 1% level. Apparently, the provision of full knowledge of the possible categories enabled *S* to orient himself toward this perceptual situation in terms of a more appropriate perceptual set.

In contrast to the above are the differences in threshold obtained for comparison of the average of the first set of four presentations as compared to the second set within each variation. None of these differences is significant at the 1% level, and only one at the 5% level. This represents some evidence for believing that a perceptual set was established by the first set of presentations, and that it was this set which was a significant determinant of each *S*'s orientation to the second set of presentations.

It is also possible to make comparisons between the three variations. Examination of the differences between the final four presentations (four categories with knowledge) for each variation discloses no significant differences between four one-digit, four two-digit, and four three-digit numbers. Moreover, examination of the differences between the first four presentations in each sequence discloses no significant differences between the 10 one-digit numbers, the 100 two-digit numbers, and the 1,000 three-digit numbers. Again these experimental results are consistent with the earlier finding of no increase in threshold for amounts of information in excess of eight categories.

Discussion

The experimental work described in this report was designed to facilitate investigation of two hypotheses: (*a*) that the threshold distance at which visual discrimination of numerals takes place is directly related to the number of alternative possibilities, and (*b*) that in a situation in which expected uncertainty exceeds actual uncertainty, the visual threshold will be a function of the number of alternatives expected by *S*, and not of the number actually possible.

With reference to the first hypothesis, significant increases in threshold values were obtained for 2 to 4 and 4 to 8 categories; decreases in threshold were also obtained for a comparison of 10 or 100 or 1,000 categories as compared to 4. This latter finding is based on the fact that significant decreases in threshold were obtained in all three variations of the second experiment for the restricted subsets with full knowledge as compared to either the total sets or the restricted subsets without knowledge to *S*.

Consideration of the categories used may suggest a possible explanation for this failure to obtain increases in thresh-

olds for conditions in excess of 8 to 10 categories. For amounts of information not in excess of ten categories, all the uncertainty was contained in the digit position. Higher amounts of information were obtained only by combining the uncertainty in this position with uncertainty in the ten's or hundred's position. Since our number system consists of only ten possibilities, larger amounts of information can be displayed only by creating sequences of any desired length out of combinations of these elements. Similarly, our language system is based upon twenty-six letters as elements and of sequences of these letters arranged in words, sentences, paragraphs, etc. One suspects that it is no accident that human communications systems are apparently formed out of a limited set of elemental symbols which can be combined into sequences according to extremely flexible rules.

To return to these experimental results, increases in threshold were obtained as long as the uncertainty in the digits position was manipulated. Once the limit of either eight or ten categories was reached, no increase in threshold was obtainable by compounding the uncertainty in the digits position with uncertainty in either of the other positions. Perhaps the perceptual decisions in each position of the display can be made independently and concurrently. If so, the magnitude of the threshold would be a function primarily of that position which contains the highest amount of uncertainty. This suggested explanation is consistent with the informal observations of some *S*s. However, further experimental work is in order if the validity of such a hypothesis is to receive confirmation.

As for the second hypothesis, the data from these experiments were consistent with the initially stated hypothesis. In the variation based upon the set of one-digit numbers, *S*s seemed to behave as if they came to expect the possibility of all one-digit numbers and to carry over this expectation to a situation in which only a restricted subset was possible. However, they appeared to be able to reorient themselves to the task in terms of the appropriate categories as soon as knowledge of this subset was made available. The findings for this variation are further corroborated by those obtained from similar variations with two- and three-digit numbers.

Summary

Two experiments were carried out to investigate the effect of the number of alternatives and the perceptual set of *S* concerning these alternatives on the visual discrimination of numerals. The visual task confronting *S*s was to discriminate accurately the particular number displayed. The threshold was defined as the maximum distance at which this discrimination was possible.

The results indicated that the difficulty of this perceptual decision varied directly with the number of possible alternatives within the limited range of not more than ten categories. A possible explanation for this finding was advanced.

In addition, variations were undertaken to measure performance under conditions in which expected uncertainty of *S*s was greater than actual uncertainty. It was shown that thresholds obtained under these conditions exceeded those obtained when full knowledge was available to *S* as to the actual set of alternatives.

References

1. Berger, C. Stroke width, form and horizontal spacing of numerals as determinants of the threshold of recognition. *J. appl. Psychol.*, 1944, **28**, 208–231, 336–346.
2. Hake, H. W., & Garner, W. R. The effect of presenting various numbers of discrete steps on scale reading accuracy. *J. exp. Psychol.*, 1951, **42**, 358–366.
3. Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.
4. Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.
5. Krulee, G. K., & Sinclair, E. J. Some behavioral implications of information theory. *Naval Res. Lab. Rep.*, 1953. Rep. No. 4119.
6. Mann, H. B. *Analysis and design of experiments.* New York: Dover, 1949.
7. Tippett, L. H. C. *The methods of statistics.* London: Williams & Norgate, 1941.

(Received for early publication February 23, 1954)